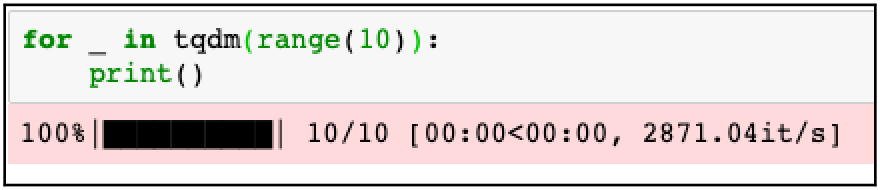
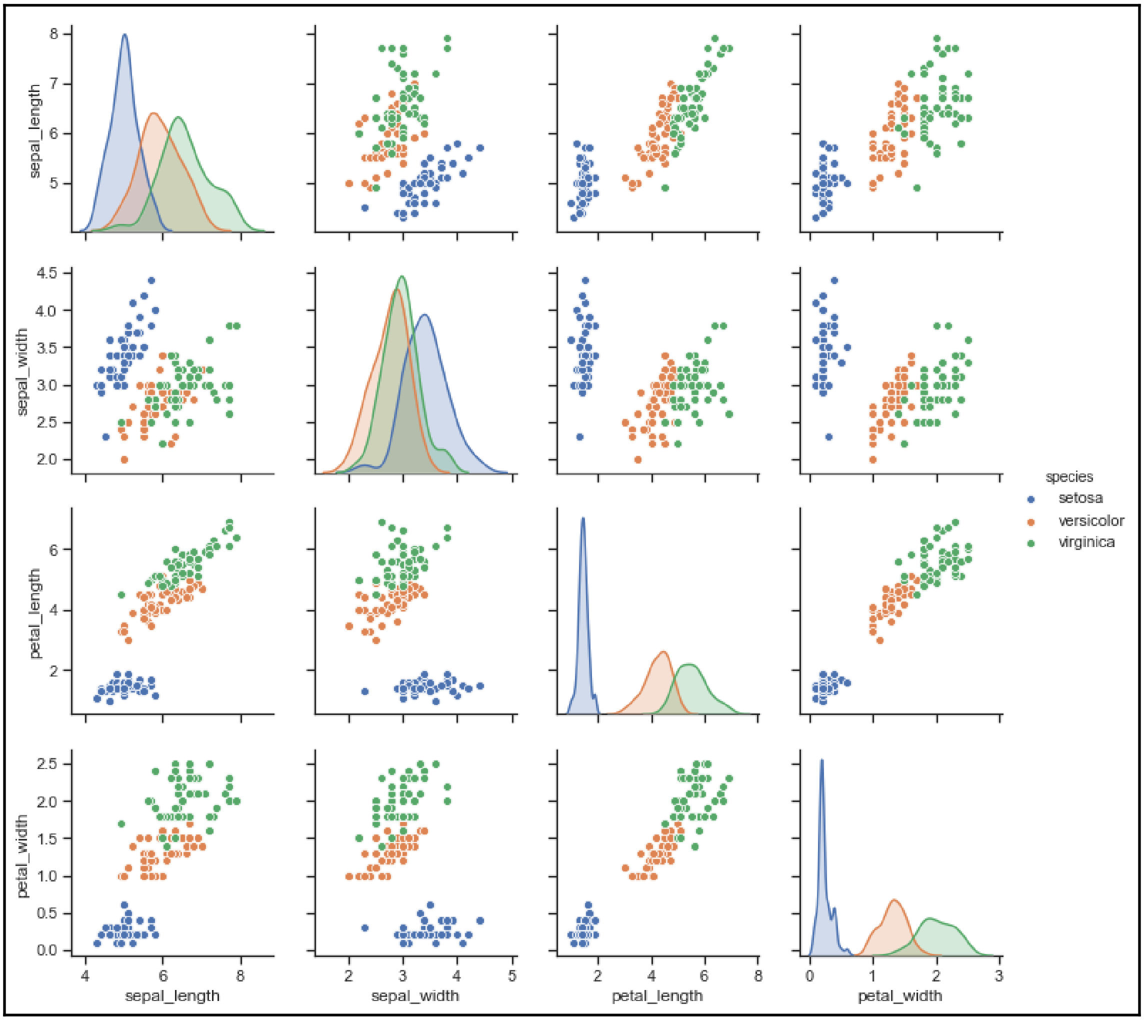
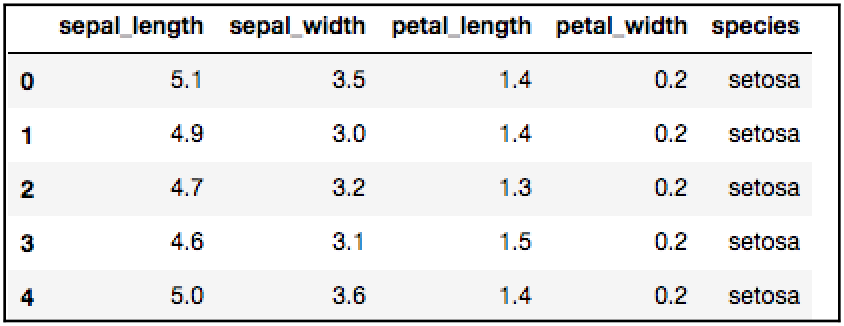
Chapter 1: Getting Started with Artificial Intelligence in Python



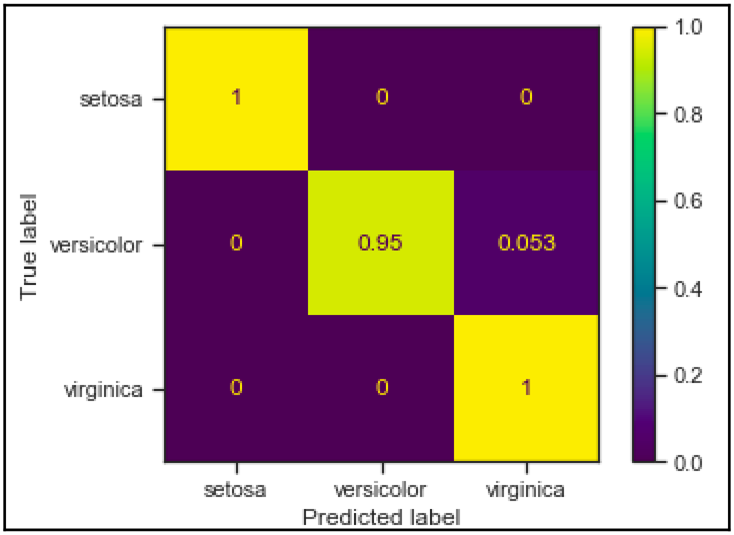
```
for _ in tqdm(range(10)):
    print()
```

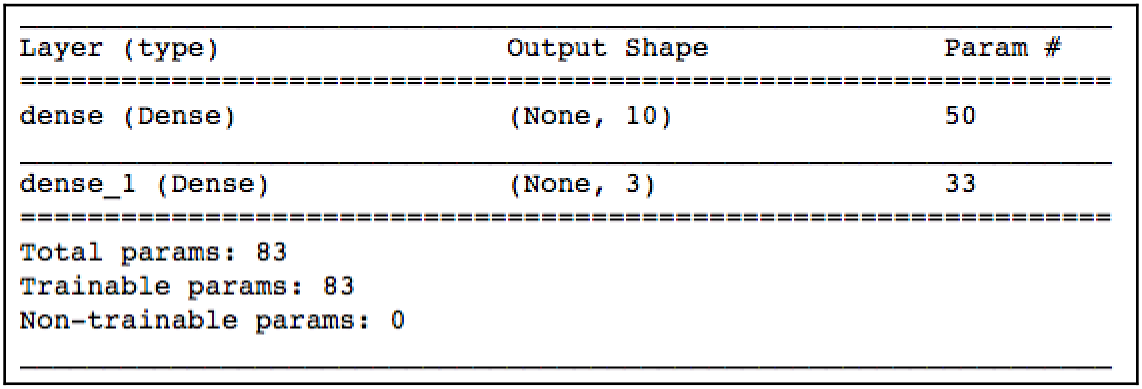
```
100%|██████████| 10/10 [00:00<00:00, 2871.04it/s]
```



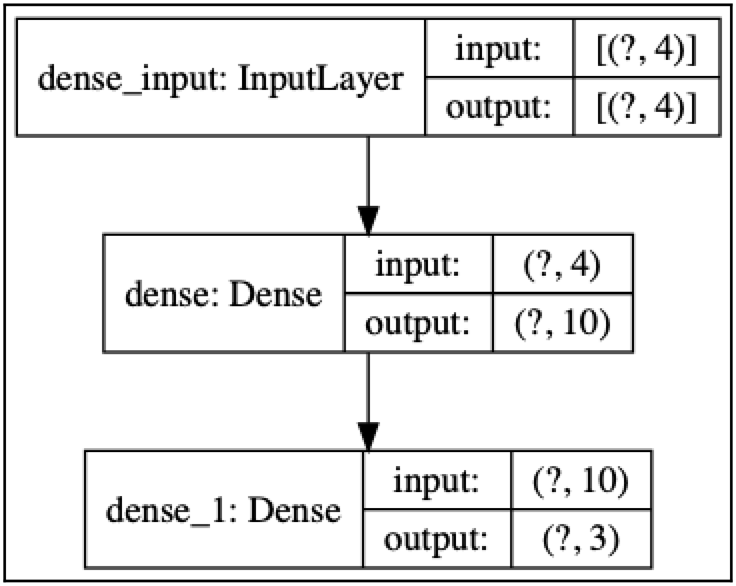


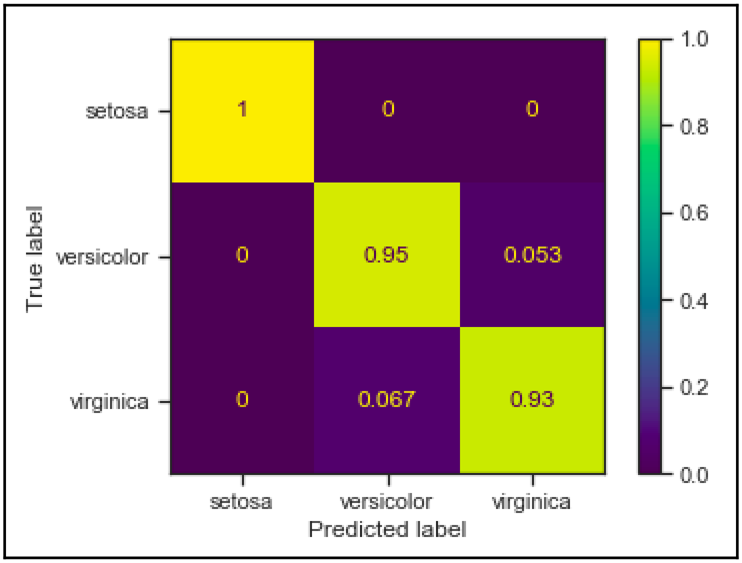
	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa



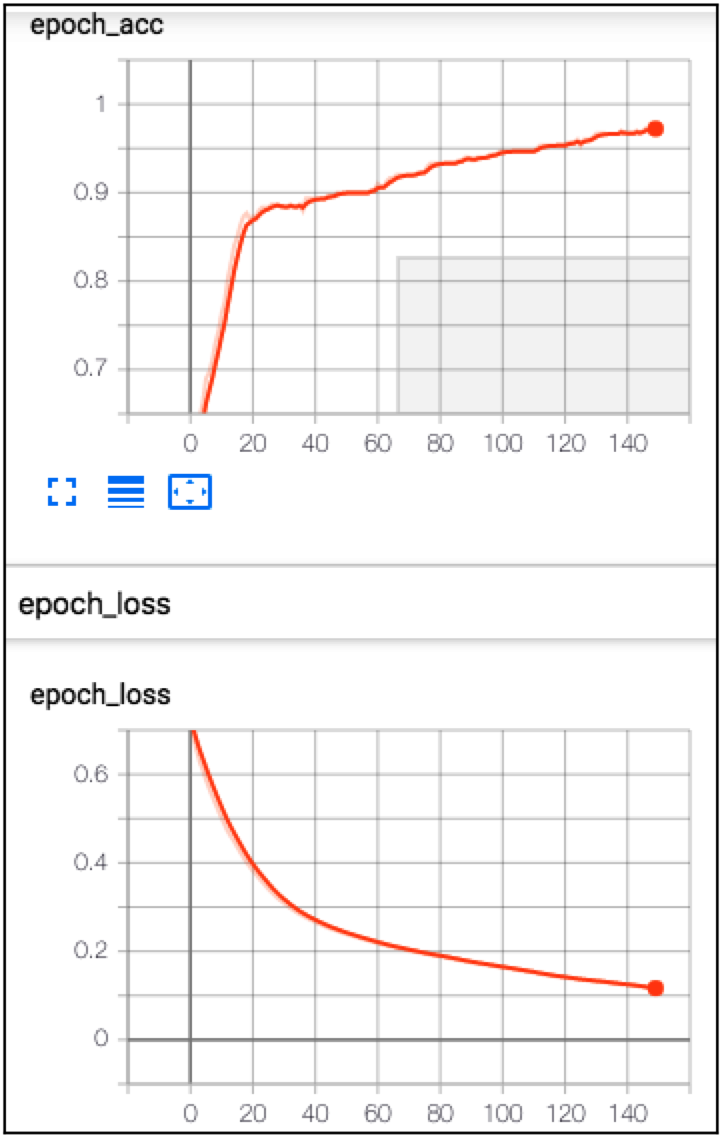


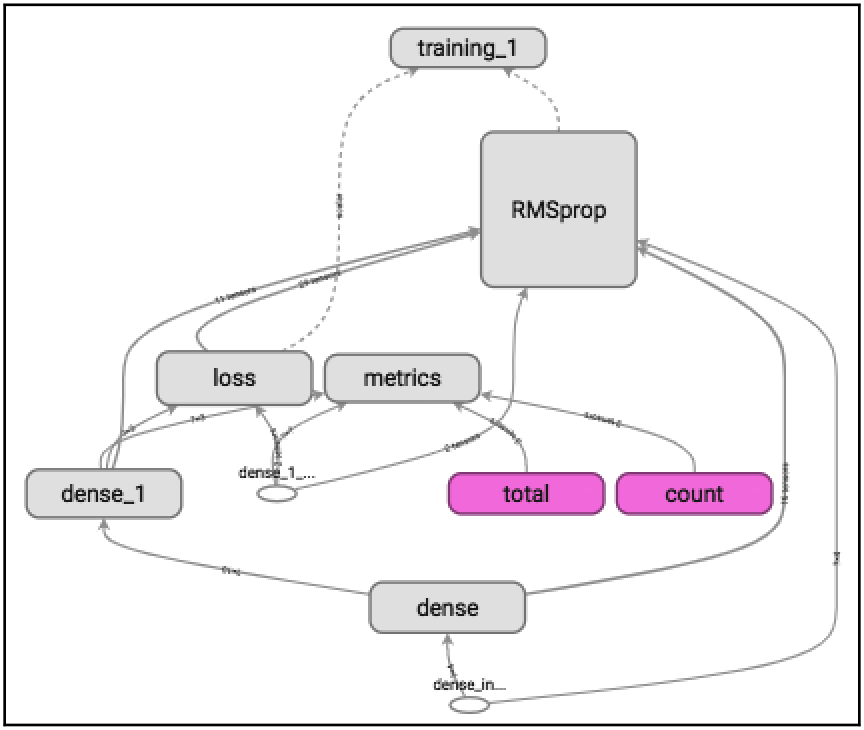
```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 10)                50

_________________________________________________________________
dense_1 (Dense)              (None, 3)                 33
=================================================================
Total params: 83
Trainable params: 83
Non-trainable params: 0

_________________________________________________________________
```

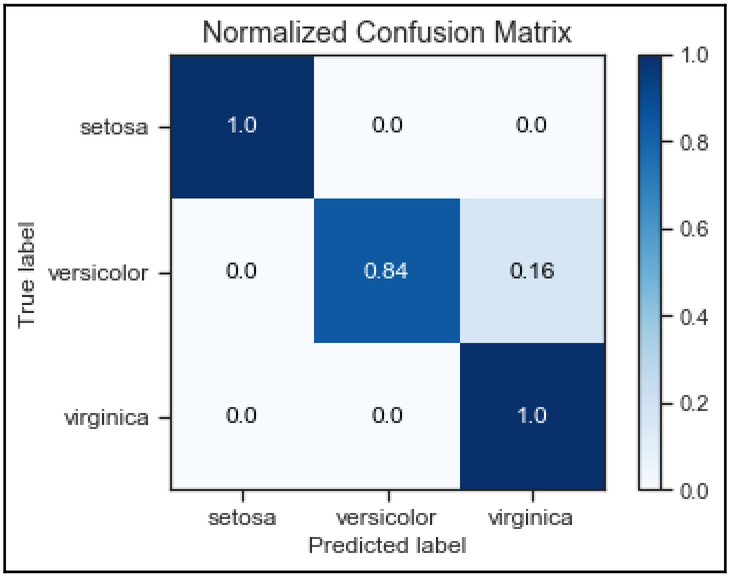




	setosa	versicolor	virginica
setosa	1	0	0
versicolor	0	0.95	0.053
virginica	0	0.067	0.93

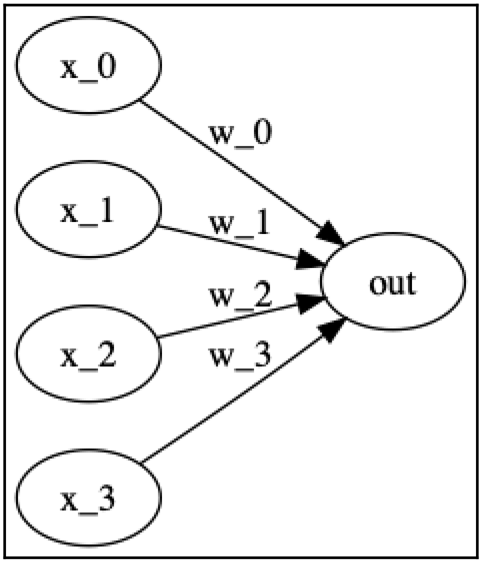

epoch_acc
1
0.9
0.8
0.7
0
20
40
60
80
100
120
140
epoch_loss
epoch_loss
0.6
0.4
0.2
0
0
20
40
60
80
100
120
140


training_1
RMSprop
loss
metrics
dense_1
dense_1_...
total
count
dense
dense_in...


Normalized Confusion Matrix
True label
setosa
versicolor
virginica
1.0
0.0
0.0
0.0
0.84
0.16
0.0
0.0
1.0
setosa
versicolor
virginica
Predicted label
1.0
0.8
0.6
0.4
0.2
0.0

$$F(x) = \sigma_m \circ f_m \circ \sigma_{m-1} \circ f_{m-1} \cdots \circ x$$

$F(x) = \sum_{i=0}^{n} x_i w_i = w \times x$, with n neurons and a single output.



$$\hat{y} = mx + b$$

$$\hat{y} = W^T x + b$$

$$b$$

$$\boldsymbol{W}$$

$$\hat{y}$$

$$y$$

$$y$$

$$\hat{y} = \begin{cases} 1 & \text{if } W^T x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$\hat{y} = \tanh(W^T x)$$

$$\mathcal{E}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2.$$

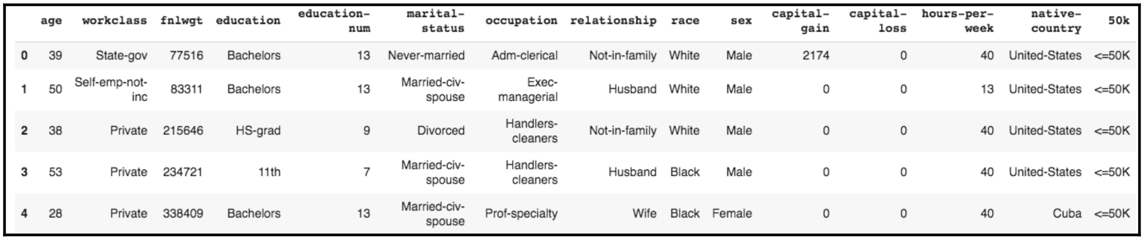
$$\Delta w_{ji}(n) = -\eta\frac{\partial\mathcal{E}(n)}{\partial v_j(n)}y_i(n)$$

$$\eta$$

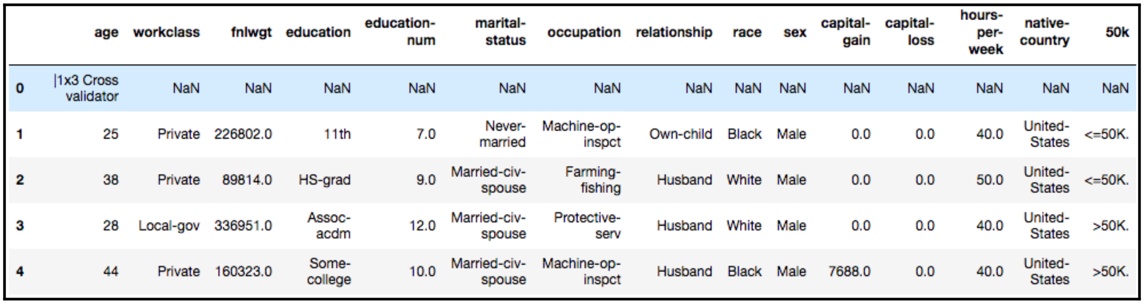
$$\mathrm{SELU}(x) = \lambda \begin{cases} x & \text{if } x > 0 \\ \alpha(\epsilon^{x} - 1) & \text{else} \end{cases}$$

$$P(y = j \mid \mathbf{x}) = \frac{e^{a_j}}{\sum_{k=1}^{K} e^{a_k}}, \text{with output activation vector } a \in \mathbb{R}^{K}.$$

$$-\sum_{c=1}^{M} y_{o,c} \log(p_{o,c})$$

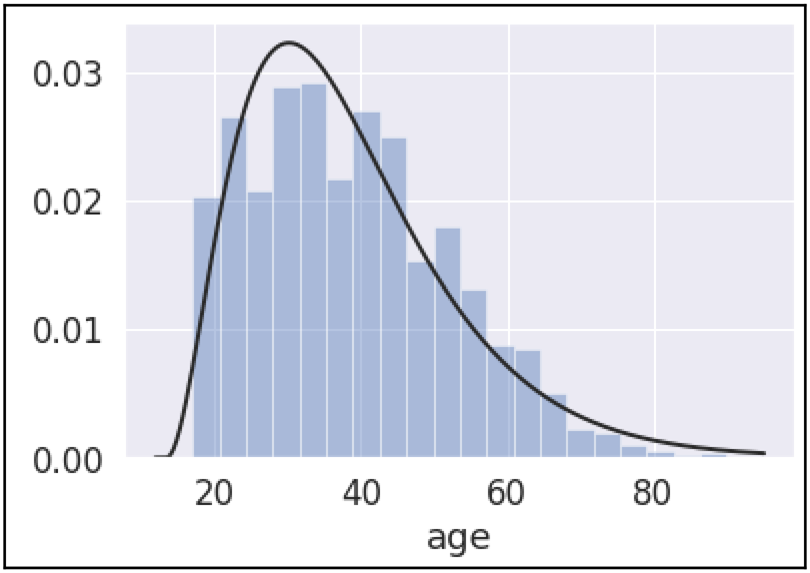


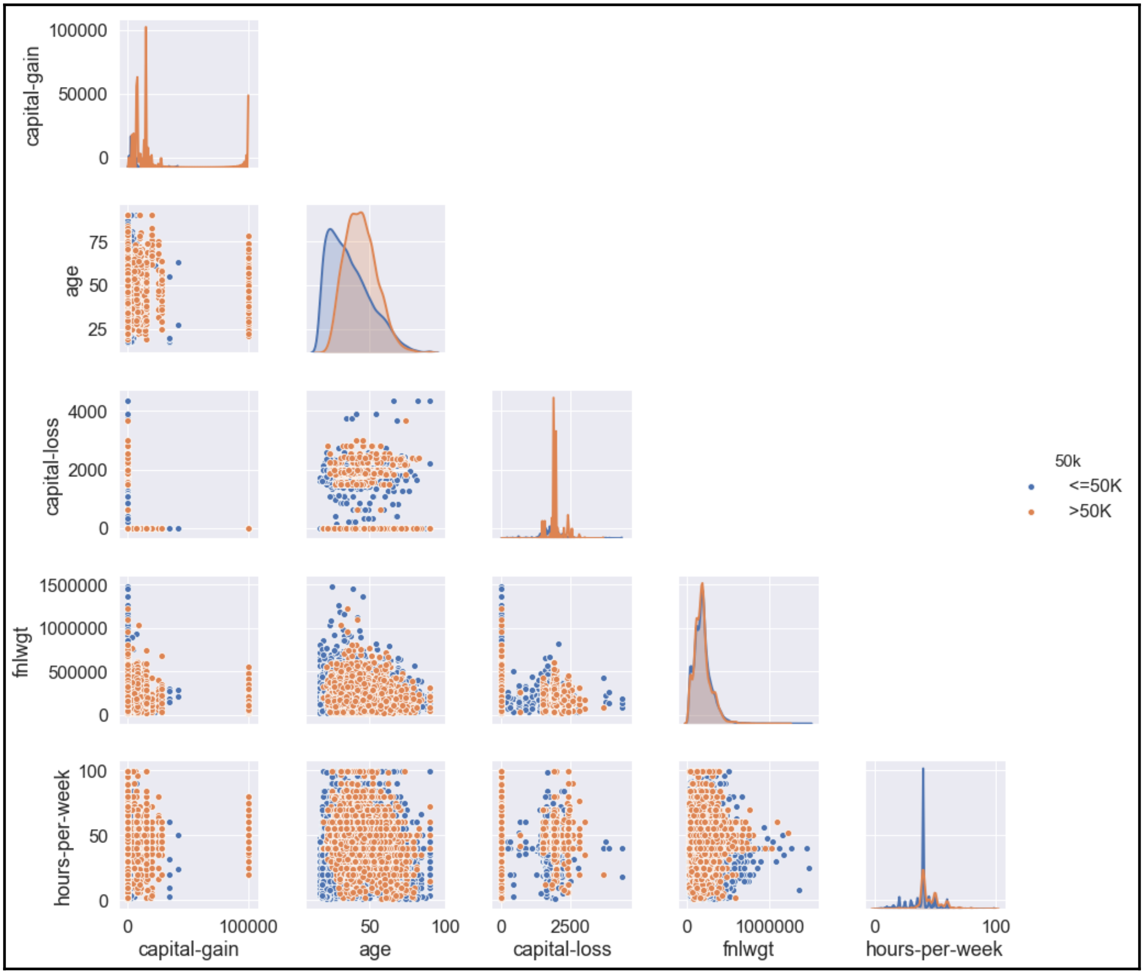
	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss	hours-per-week	native-country	50k
0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States	<=50K
1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States	<=50K
2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States	<=50K
3	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States	<=50K
4	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	<=50K

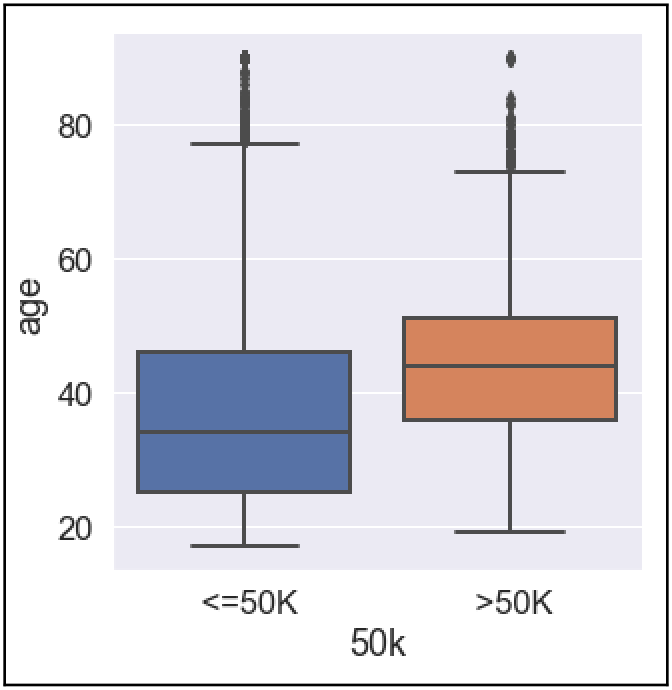


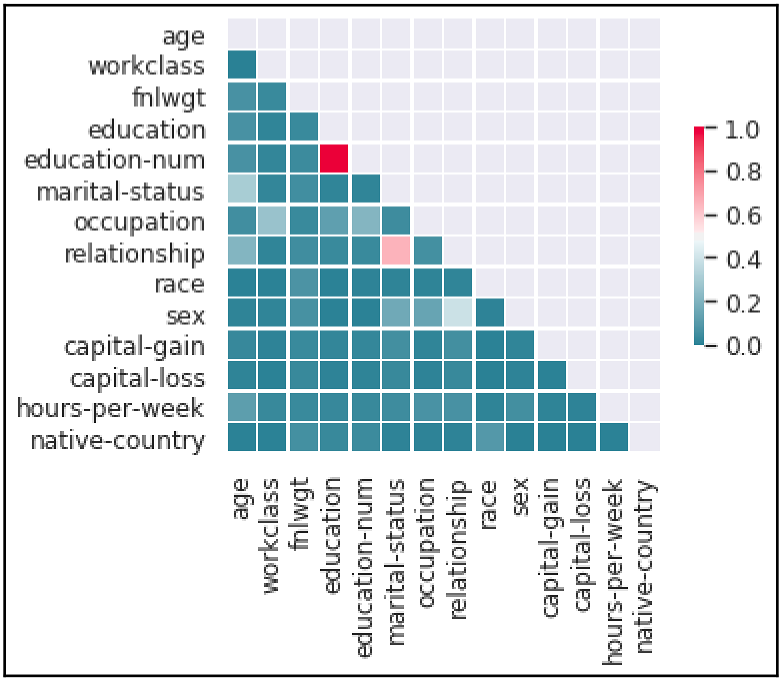
	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss	hours-per-week	native-country	50k
0	|1x3 Cross validator	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	25	Private	226802.0	11th	7.0	Never-married	Machine-op-inspct	Own-child	Black	Male	0.0	0.0	40.0	United-States	<=50K.
2	38	Private	89814.0	HS-grad	9.0	Married-civ-spouse	Farming-fishing	Husband	White	Male	0.0	0.0	50.0	United-States	<=50K.
3	28	Local-gov	336951.0	Assoc-acdm	12.0	Married-civ-spouse	Protective-serv	Husband	White	Male	0.0	0.0	40.0	United-States	>50K.
4	44	Private	160323.0	Some-college	10.0	Married-civ-spouse	Machine-op-inspct	Husband	Black	Male	7688.0	0.0	40.0	United-States	>50K.

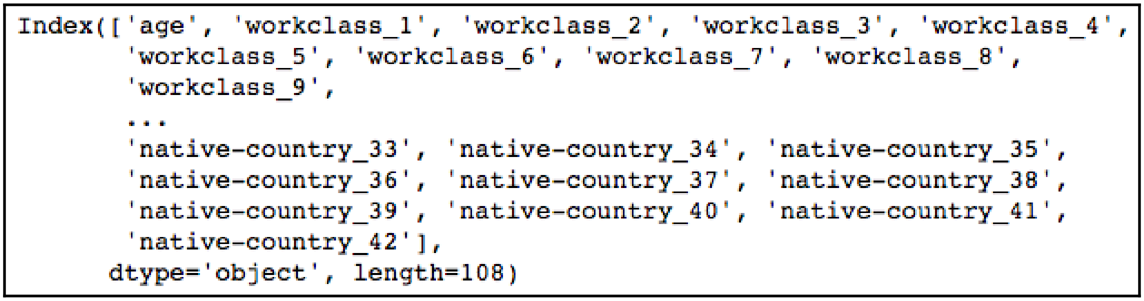
	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss	hours-per-week	native-country
0	39	1	77516	1	13	1	1	1	1	1	2174	0	40	1
1	50	2	83311	1	13	2	2	2	1	1	0	0	13	1
2	38	3	215646	2	9	3	3	1	1	1	0	0	40	1
3	53	3	234721	3	7	2	3	2	2	1	0	0	40	1
4	28	3	338409	1	13	2	4	3	2	2	0	0	40	2



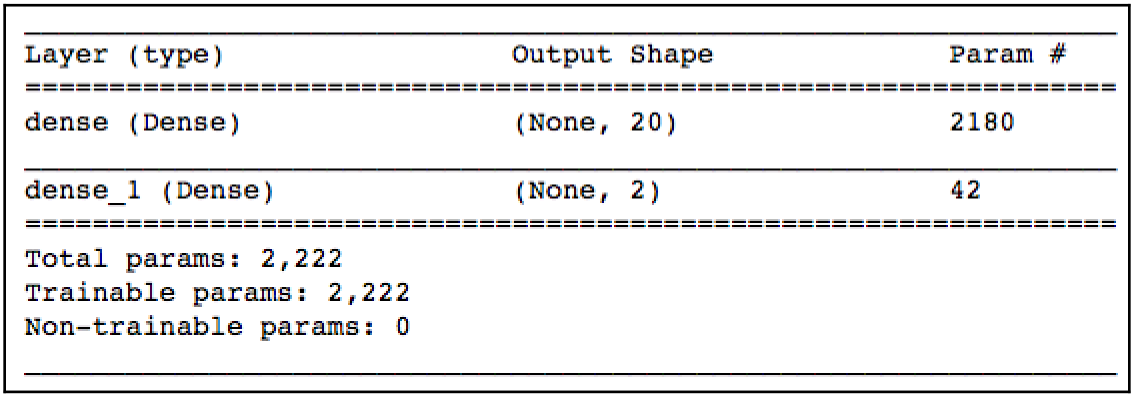

capital-gain
100000
50000
0
age
75
50
25
capital-loss
4000
2000
0
fnlwgt
1500000
1000000
500000
0
hours-per-week
100
50
0
0
100000
capital-gain
50
100
age
0
2500
capital-loss
0
1000000
fnlwgt
0
100
hours-per-week
50k
<=50K
>50K


80
60
40
20
age
<=50K
>50K
50k

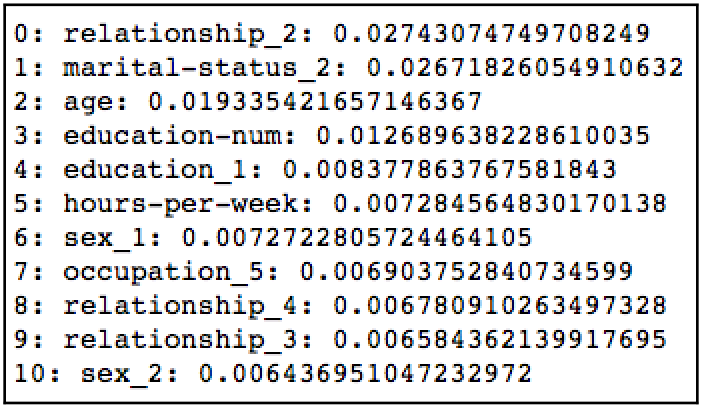

age
workclass
fnlwgt
education
education-num
marital-status
occupation
relationship
race
sex
capital-gain
capital-loss
hours-per-week
native-country
1.0
0.8
0.6
0.4
0.2
0.0



```
Index(['age', 'workclass_1', 'workclass_2', 'workclass_3', 'workclass_4',
       'workclass_5', 'workclass_6', 'workclass_7', 'workclass_8',
       'workclass_9',
       ...
       'native-country_33', 'native-country_34', 'native-country_35',
       'native-country_36', 'native-country_37', 'native-country_38',
       'native-country_39', 'native-country_40', 'native-country_41',
       'native-country_42'],
      dtype='object', length=108)
```

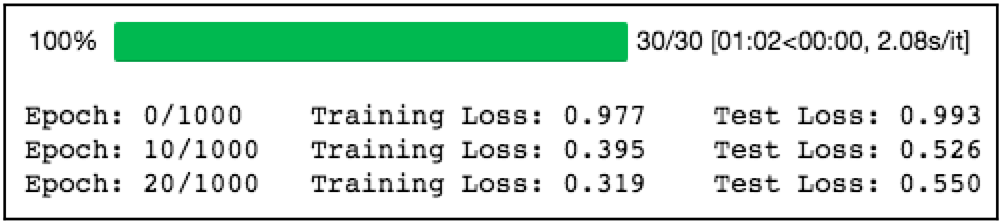


```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 20)                2180

_________________________________________________________________
dense_1 (Dense)              (None, 2)                 42
=================================================================
Total params: 2,222
Trainable params: 2,222
Non-trainable params: 0

_________________________________________________________________
```

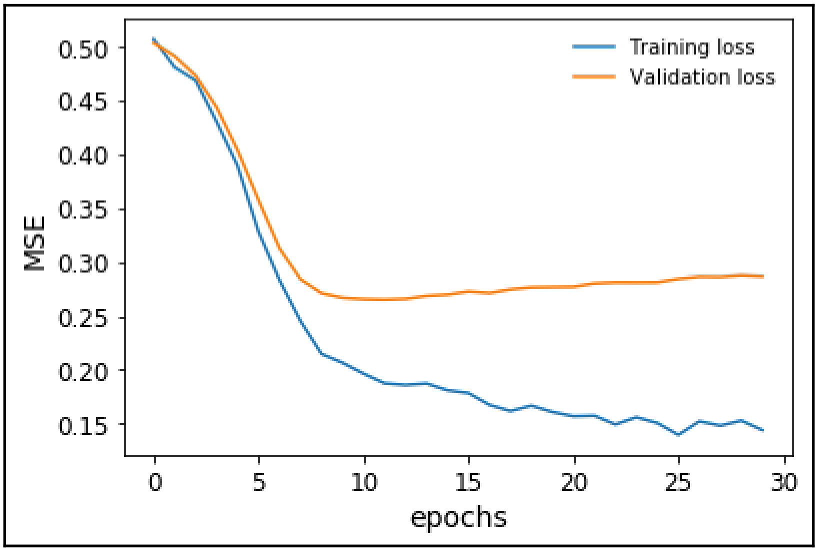


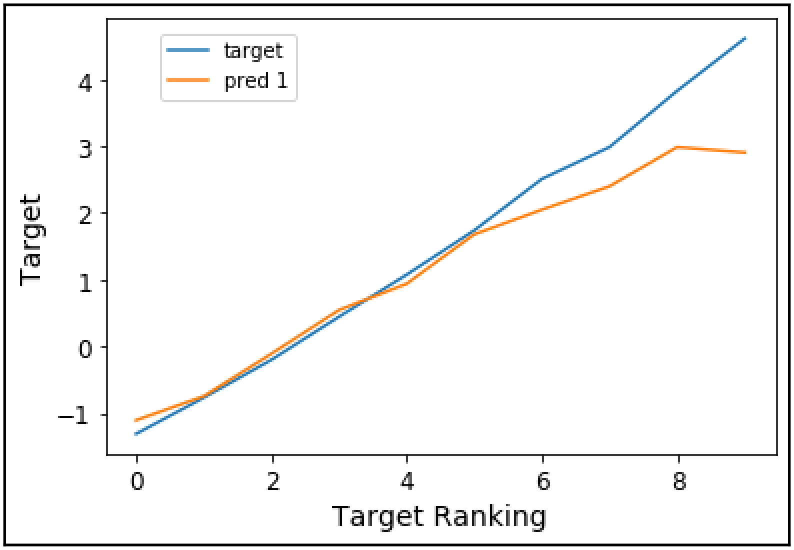
```
0: relationship_2: 0.02743074749708249
1: marital-status_2: 0.02671826054910632
2: age: 0.019335421657146367
3: education-num: 0.012689638228610035
4: education_1: 0.008377863767581843
5: hours-per-week: 0.007284564830170138
6: sex_1: 0.0072722805724464105
7: occupation_5: 0.006903752840734599
8: relationship_4: 0.006780910263497328
9: relationship_3: 0.006584362139917695
10: sex_2: 0.006436951047232972
```

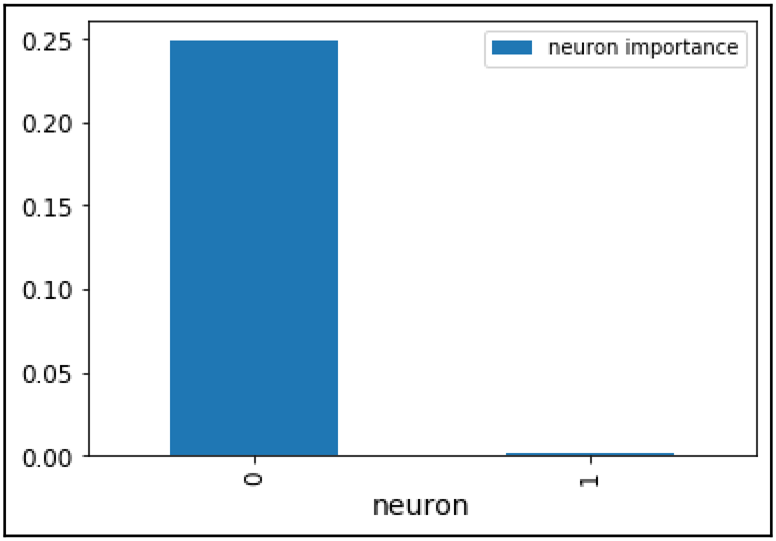
$$\mathrm{len}(\mathrm{data})/\mathrm{batch\_size}$$

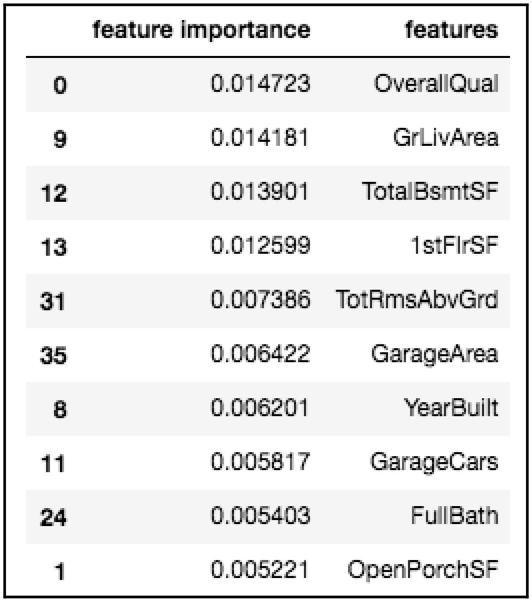


```
100%                                              30/30 [01:02<00:00, 2.08s/it]

Epoch: 0/1000     Training Loss: 0.977     Test Loss: 0.993
Epoch: 10/1000    Training Loss: 0.395     Test Loss: 0.526
Epoch: 20/1000    Training Loss: 0.319     Test Loss: 0.550
```

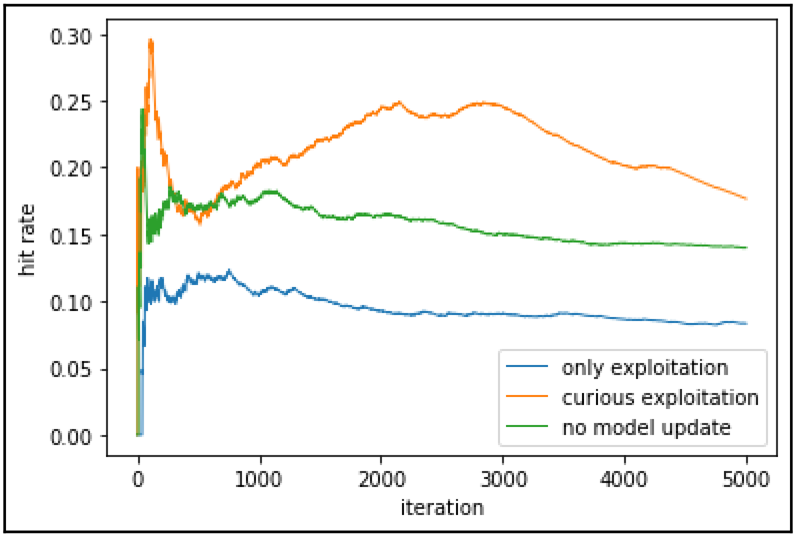



target
pred 1
Target
4
3
2
1
0
−1
0
2
4
6
8
Target Ranking


neuron importance
0.25
0.20
0.15
0.10
0.05
0.00
0
1
neuron



	feature importance	features
0	0.014723	OverallQual
9	0.014181	GrLivArea
12	0.013901	TotalBsmtSF
13	0.012599	1stFlrSF
31	0.007386	TotRmsAbvGrd
35	0.006422	GarageArea
8	0.006201	YearBuilt
11	0.005817	GarageCars
24	0.005403	FullBath
1	0.005221	OpenPorchSF



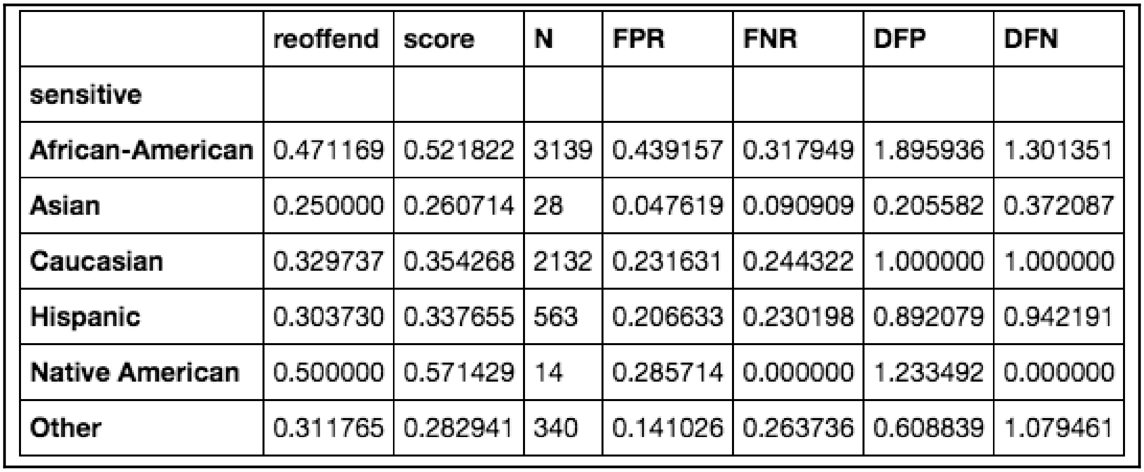
if $G(A_i, D) > G(A_j, D) + \epsilon(|S|, \delta), \forall j \neq i,$

For dataset D and attribute A_i, and gain function $G()$,

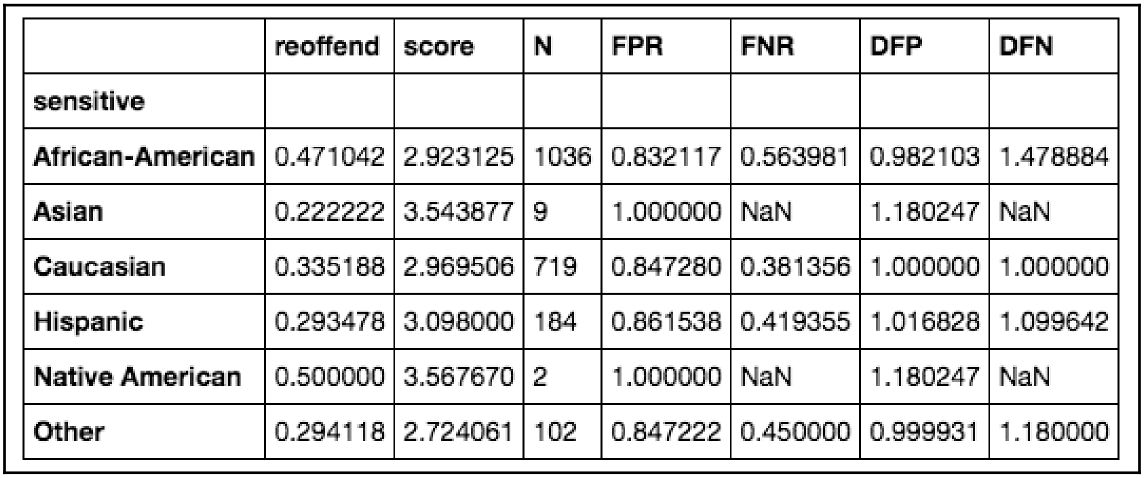
then we conclude that A_i is the best attribute with probability $1 - \delta$.

$$\frac{\text{n\_samples}}{\text{n\_classes} \times \text{np.bincount}(y)}$$

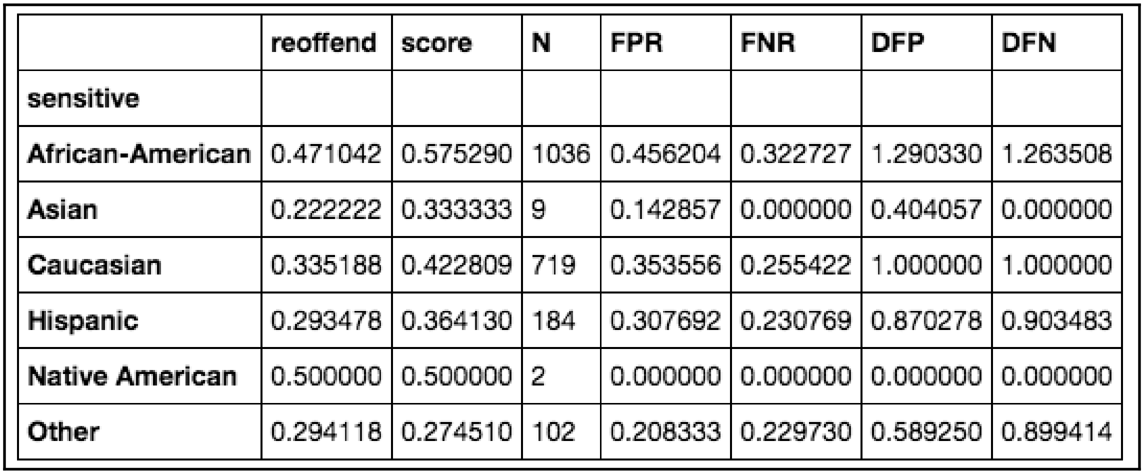
$$\text{AIR}_{\text{metric}} = metric_{\text{protected\_group}} / metric_{\text{norm\_group}}$$



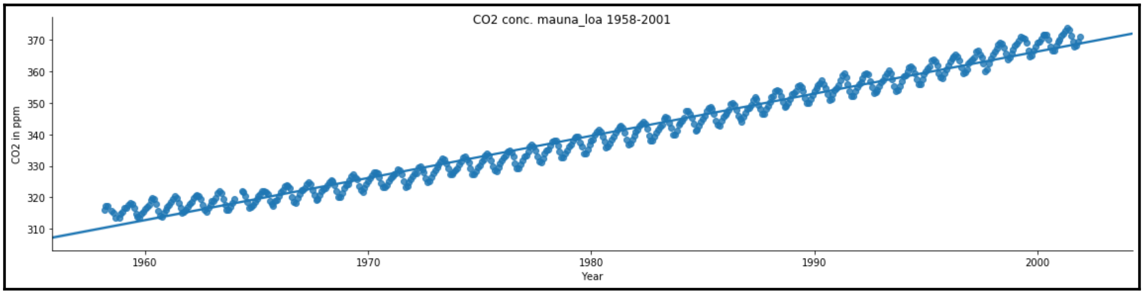
	reoffend	score	N	FPR	FNR	DFP	DFN
sensitive							
African-American	0.471169	0.521822	3139	0.439157	0.317949	1.895936	1.301351
Asian	0.250000	0.260714	28	0.047619	0.090909	0.205582	0.372087
Caucasian	0.329737	0.354268	2132	0.231631	0.244322	1.000000	1.000000
Hispanic	0.303730	0.337655	563	0.206633	0.230198	0.892079	0.942191
Native American	0.500000	0.571429	14	0.285714	0.000000	1.233492	0.000000
Other	0.311765	0.282941	340	0.141026	0.263736	0.608839	1.079461

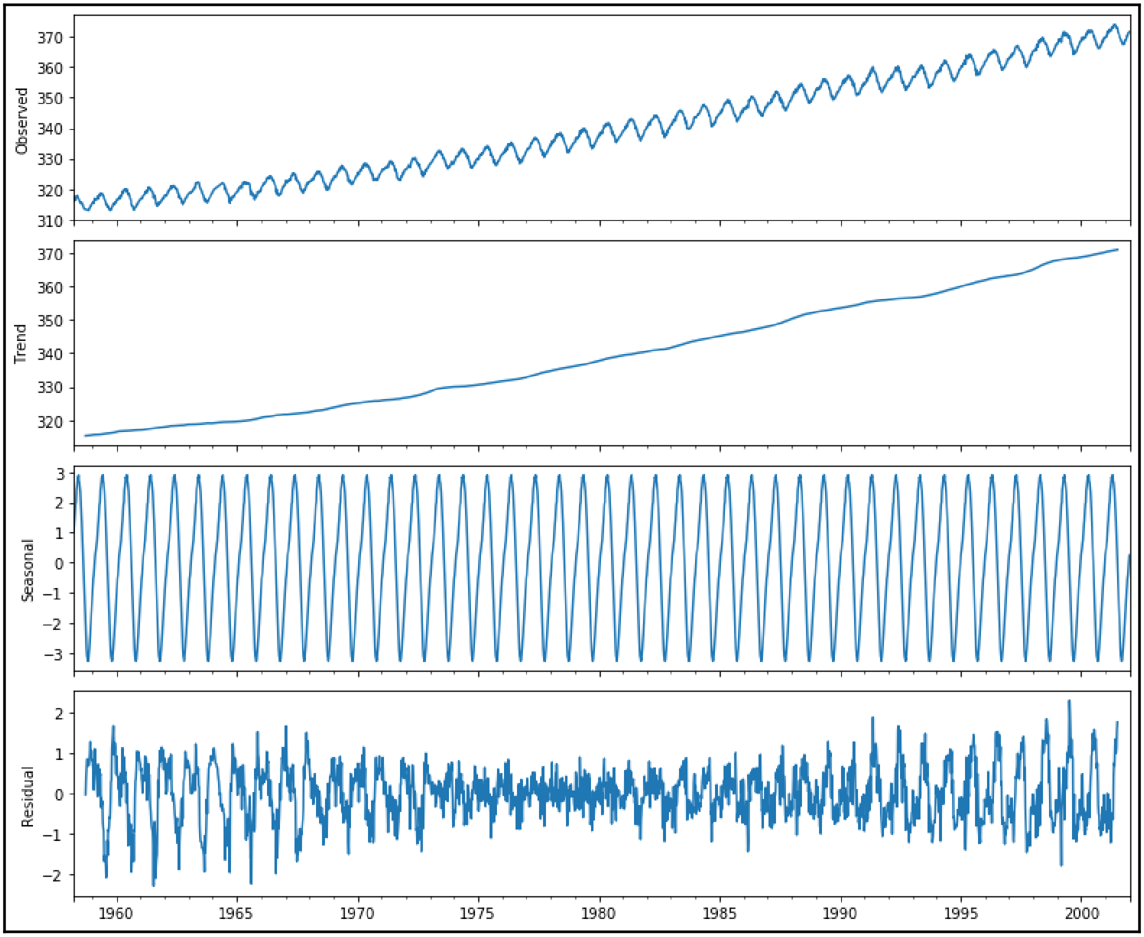


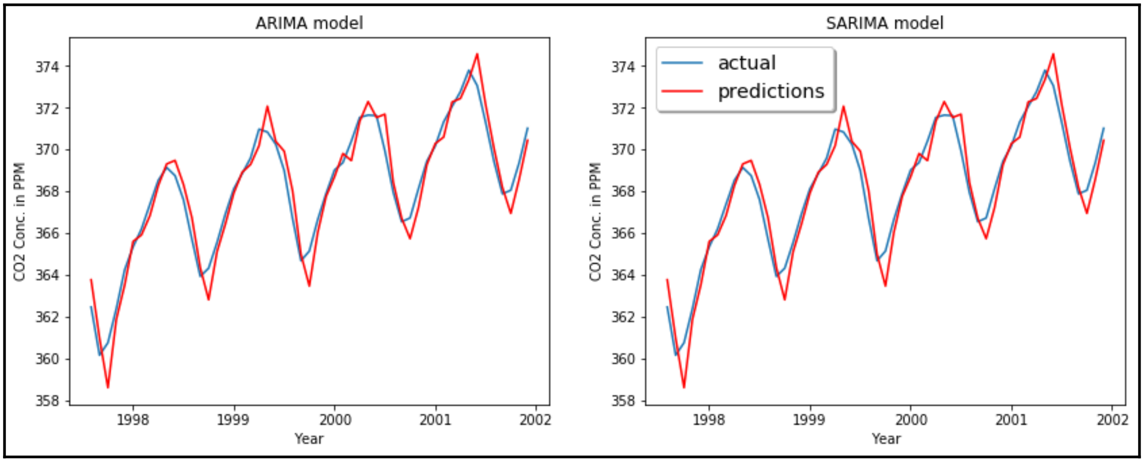
	reoffend	score	N	FPR	FNR	DFP	DFN
sensitive							
African-American	0.471042	2.923125	1036	0.832117	0.563981	0.982103	1.478884
Asian	0.222222	3.543877	9	1.000000	NaN	1.180247	NaN
Caucasian	0.335188	2.969506	719	0.847280	0.381356	1.000000	1.000000
Hispanic	0.293478	3.098000	184	0.861538	0.419355	1.016828	1.099642
Native American	0.500000	3.567670	2	1.000000	NaN	1.180247	NaN
Other	0.294118	2.724061	102	0.847222	0.450000	0.999931	1.180000



	reoffend	score	N	FPR	FNR	DFP	DFN
sensitive							
African-American	0.471042	0.575290	1036	0.456204	0.322727	1.290330	1.263508
Asian	0.222222	0.333333	9	0.142857	0.000000	0.404057	0.000000
Caucasian	0.335188	0.422809	719	0.353556	0.255422	1.000000	1.000000
Hispanic	0.293478	0.364130	184	0.307692	0.230769	0.870278	0.903483
Native American	0.500000	0.500000	2	0.000000	0.000000	0.000000	0.000000
Other	0.294118	0.274510	102	0.208333	0.229730	0.589250	0.899414




Observed
370
360
350
340
330
320
310
Trend
370
360
350
340
330
320
Seasonal
3
2
1
0
−1
−2
−3
Residual
2
1
0
−1
−2
1960
1965
1970
1975
1980
1985
1990
1995
2000



$$0, \ldots, N$$

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t \,,$$

$$\varphi_1 , \ldots , \varphi_p$$

$$c$$

$$\varepsilon_t$$

$$p$$

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i},$$

$$q$$

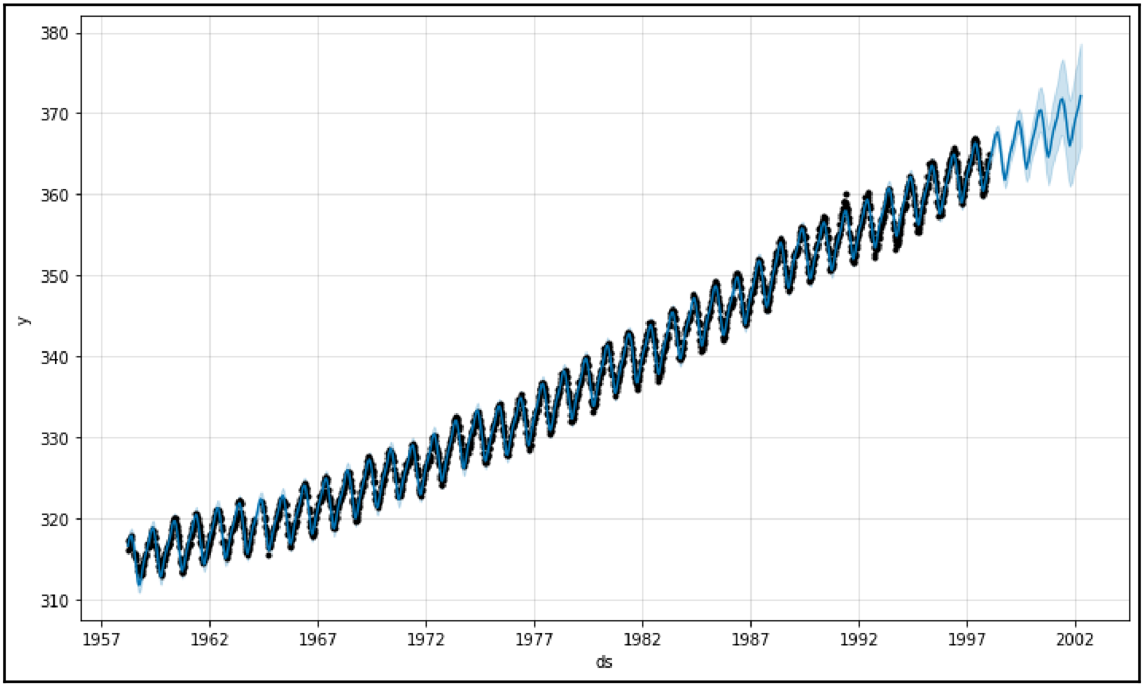
$$\theta_1, \ldots, \theta_q$$

$$\mu$$

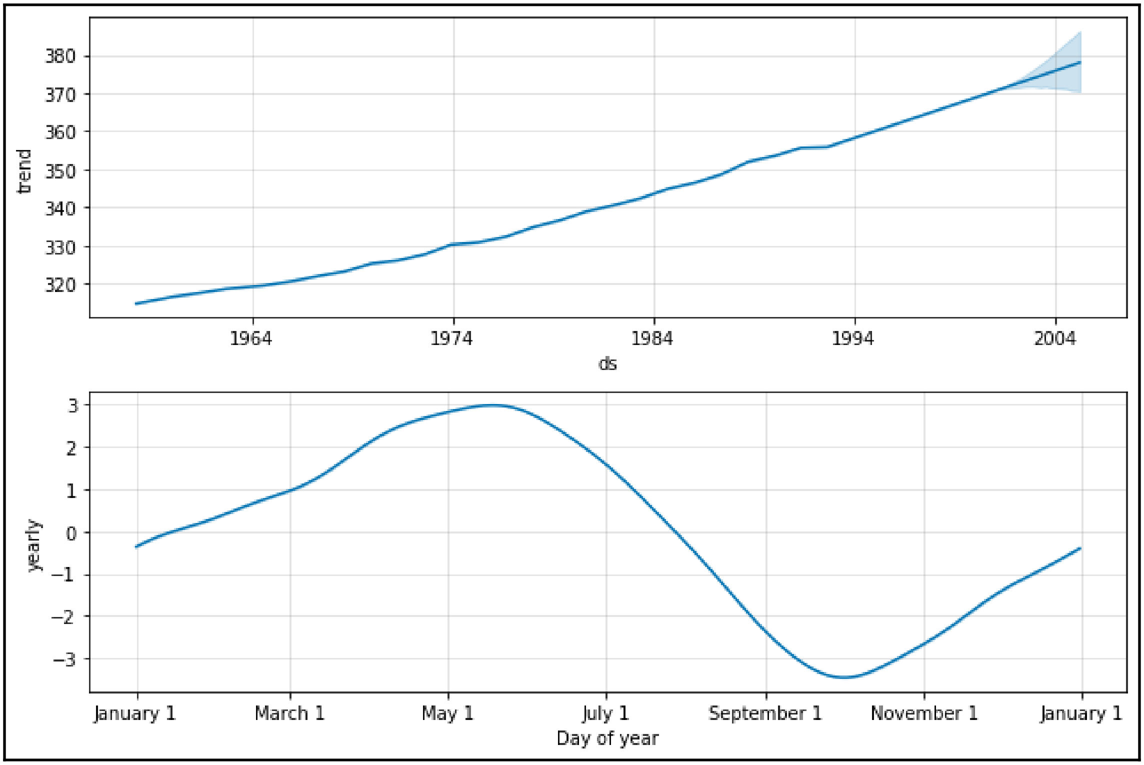
$$X$$

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}.$$

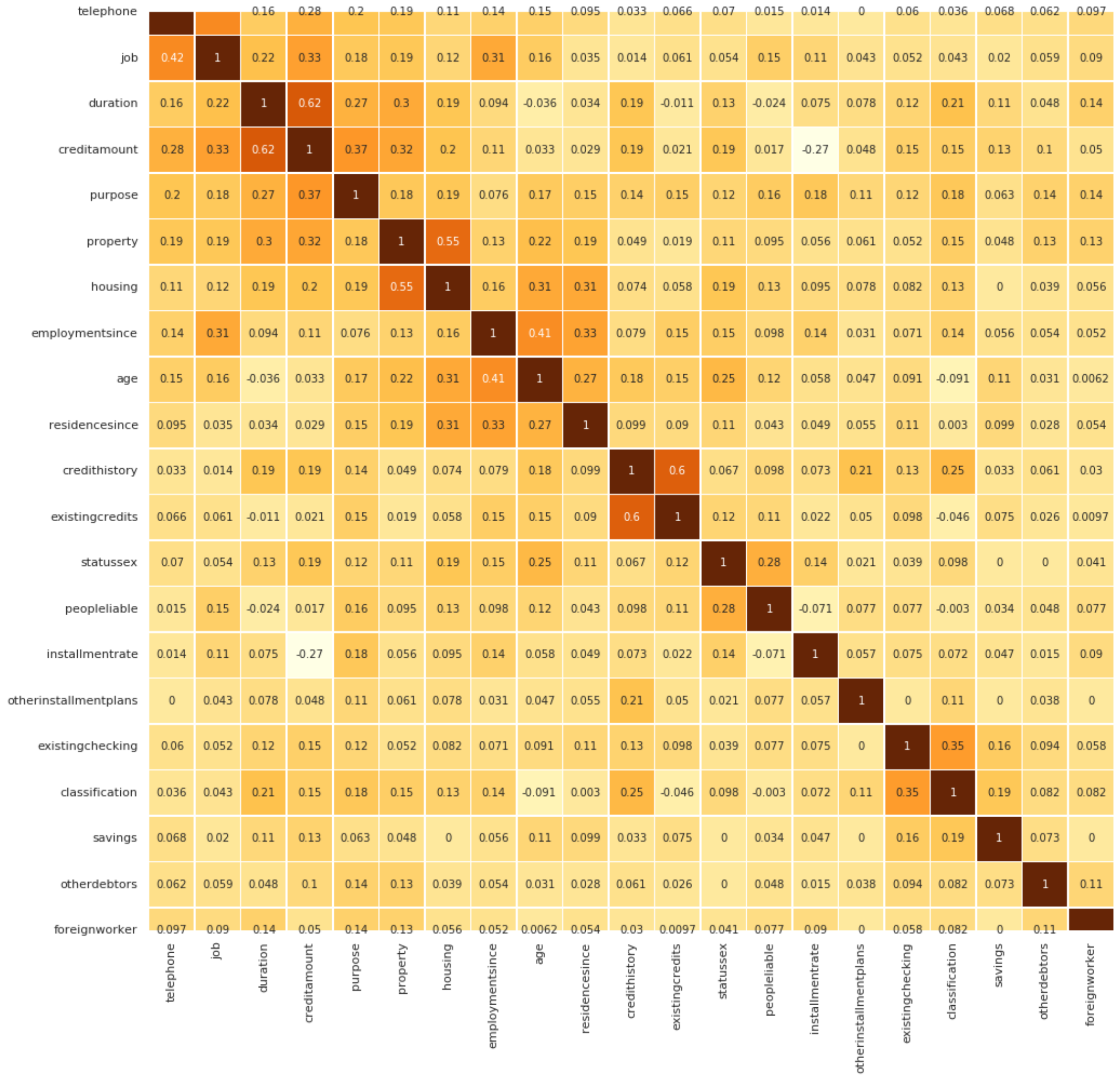
$$x'_t = x_t - x_{t-1}.$$



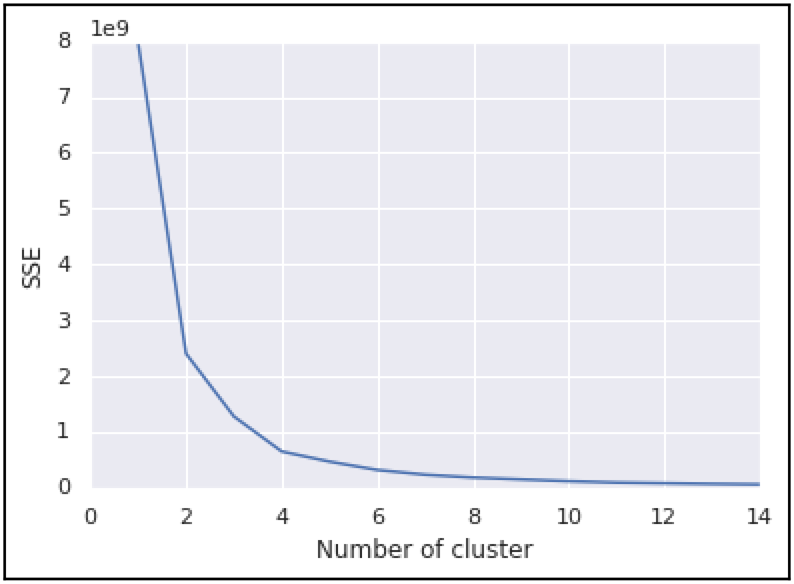
380
370
360
350
340
330
320
310
y
1957
1962
1967
1972
1977
1982
1987
1992
1997
2002
ds

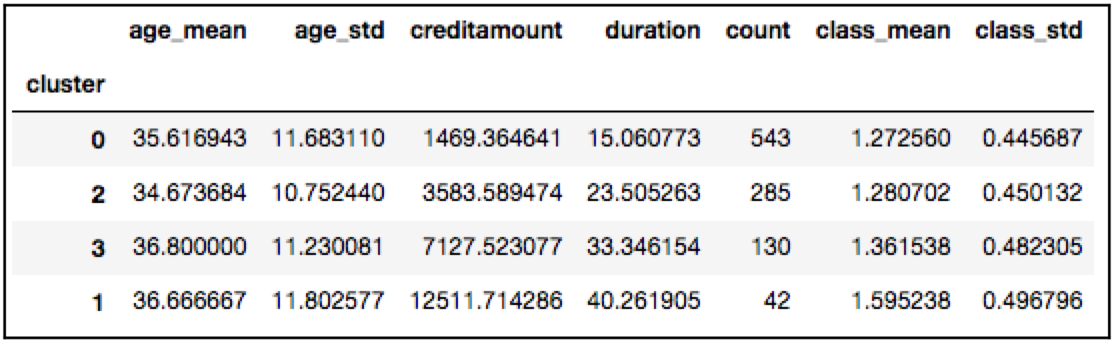

380
370
360
350
340
330
320
trend
1964
1974
1984
1994
2004
ds
3
2
1
0
−1
−2
−3
yearly
January 1
March 1
May 1
July 1
September 1
November 1
January 1
Day of year

Chapter 3: Patterns, Outliers, and Recommendations

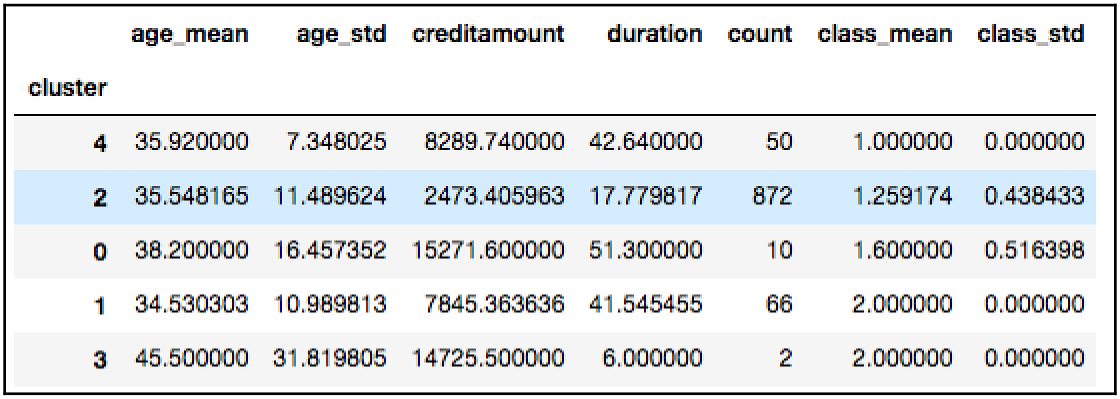


	telephone	job	duration	creditamount	purpose	property	housing	employmentsince	age	residencesince	credithistory	existingcredits	statussex	peopleliable	installmentrate	otherinstallmentplans	existingchecking	classification	savings	otherdebtors	foreignworker
telephone			0.16	0.28	0.2	0.19	0.11	0.14	0.15	0.095	0.033	0.066	0.07	0.015	0.014	0	0.06	0.036	0.068	0.062	0.097
job	0.42	1	0.22	0.33	0.18	0.19	0.12	0.31	0.16	0.035	0.014	0.061	0.054	0.15	0.11	0.043	0.052	0.043	0.02	0.059	0.09
duration	0.16	0.22	1	0.62	0.27	0.3	0.19	0.094	-0.036	0.034	0.19	-0.011	0.13	-0.024	0.075	0.078	0.12	0.21	0.11	0.048	0.14
creditamount	0.28	0.33	0.62	1	0.37	0.32	0.2	0.11	0.033	0.029	0.19	0.021	0.19	0.017	-0.27	0.048	0.15	0.15	0.13	0.1	0.05
purpose	0.2	0.18	0.27	0.37	1	0.18	0.19	0.076	0.17	0.15	0.14	0.15	0.12	0.16	0.18	0.11	0.12	0.18	0.063	0.14	0.14
property	0.19	0.19	0.3	0.32	0.18	1	0.55	0.13	0.22	0.19	0.049	0.019	0.11	0.095	0.056	0.061	0.052	0.15	0.048	0.13	0.13
housing	0.11	0.12	0.19	0.2	0.19	0.55	1	0.16	0.31	0.31	0.074	0.058	0.19	0.13	0.095	0.078	0.082	0.13	0	0.039	0.056
employmentsince	0.14	0.31	0.094	0.11	0.076	0.13	0.16	1	0.41	0.33	0.079	0.15	0.15	0.098	0.14	0.031	0.071	0.14	0.056	0.054	0.052
age	0.15	0.16	-0.036	0.033	0.17	0.22	0.31	0.41	1	0.27	0.18	0.15	0.25	0.12	0.058	0.047	0.091	-0.091	0.11	0.031	0.0062
residencesince	0.095	0.035	0.034	0.029	0.15	0.19	0.31	0.33	0.27	1	0.099	0.09	0.11	0.043	0.049	0.055	0.11	0.003	0.099	0.028	0.054
credithistory	0.033	0.014	0.19	0.19	0.14	0.049	0.074	0.079	0.18	0.099	1	0.6	0.067	0.098	0.073	0.21	0.13	0.25	0.033	0.061	0.03
existingcredits	0.066	0.061	-0.011	0.021	0.15	0.019	0.058	0.15	0.15	0.09	0.6	1	0.12	0.11	0.022	0.05	0.098	-0.046	0.075	0.026	0.0097
statussex	0.07	0.054	0.13	0.19	0.12	0.11	0.19	0.15	0.25	0.11	0.067	0.12	1	0.28	0.14	0.021	0.039	0.098	0	0	0.041
peopleliable	0.015	0.15	-0.024	0.017	0.16	0.095	0.13	0.098	0.12	0.043	0.098	0.11	0.28	1	-0.071	0.077	0.077	-0.003	0.034	0.048	0.077
installmentrate	0.014	0.11	0.075	-0.27	0.18	0.056	0.095	0.14	0.058	0.049	0.073	0.022	0.14	-0.071	1	0.057	0.075	0.072	0.047	0.015	0.09
otherinstallmentplans	0	0.043	0.078	0.048	0.11	0.061	0.078	0.031	0.047	0.055	0.21	0.05	0.021	0.077	0.057	1	0	0.11	0	0.038	0
existingchecking	0.06	0.052	0.12	0.15	0.12	0.052	0.082	0.071	0.091	0.11	0.13	0.098	0.039	0.077	0.075	0	1	0.35	0.16	0.094	0.058
classification	0.036	0.043	0.21	0.15	0.18	0.15	0.13	0.14	-0.091	0.003	0.25	-0.046	0.098	-0.003	0.072	0.11	0.35	1	0.19	0.082	0.082
savings	0.068	0.02	0.11	0.13	0.063	0.048	0	0.056	0.11	0.099	0.033	0.075	0	0.034	0.047	0	0.16	0.19	1	0.073	0
otherdebtors	0.062	0.059	0.048	0.1	0.14	0.13	0.039	0.054	0.031	0.028	0.061	0.026	0	0.048	0.015	0.038	0.094	0.082	0.073	1	0.11
foreignworker	0.097	0.09	0.14	0.05	0.14	0.13	0.056	0.052	0.0062	0.054	0.03	0.0097	0.041	0.077	0.09	0	0.058	0.082	0	0.11	



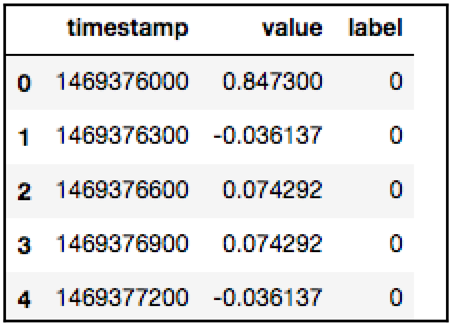


	age_mean	age_std	creditamount	duration	count	class_mean	class_std
cluster							
0	35.616943	11.683110	1469.364641	15.060773	543	1.272560	0.445687
2	34.673684	10.752440	3583.589474	23.505263	285	1.280702	0.450132
3	36.800000	11.230081	7127.523077	33.346154	130	1.361538	0.482305
1	36.666667	11.802577	12511.714286	40.261905	42	1.595238	0.496796

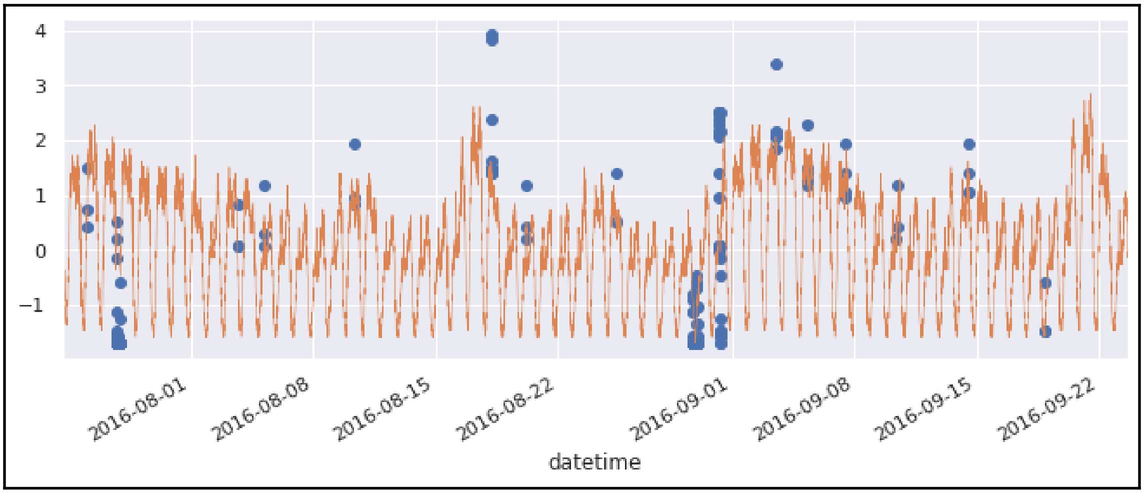


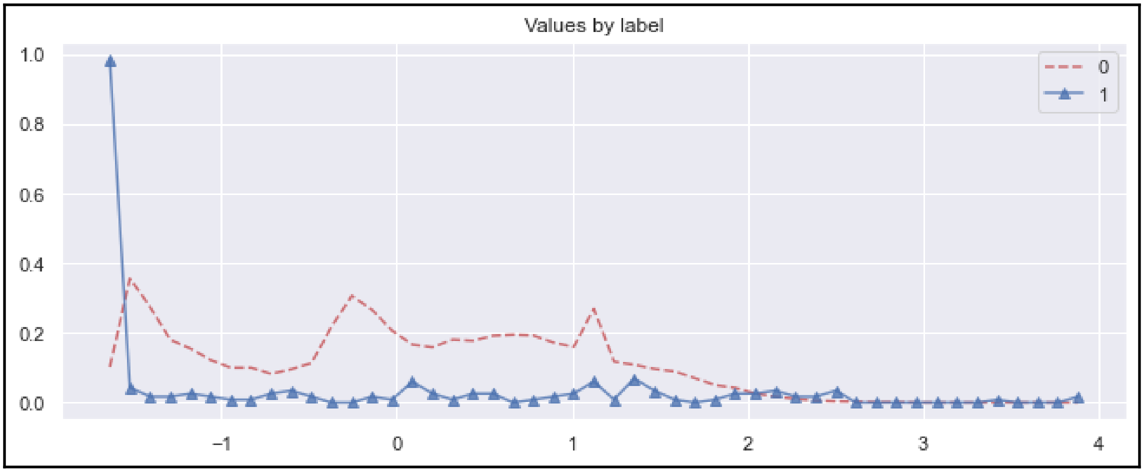
	age_mean	age_std	creditamount	duration	count	class_mean	class_std
cluster							
4	35.920000	7.348025	8289.740000	42.640000	50	1.000000	0.000000
2	35.548165	11.489624	2473.405963	17.779817	872	1.259174	0.438433
0	38.200000	16.457352	15271.600000	51.300000	10	1.600000	0.516398
1	34.530303	10.989813	7845.363636	41.545455	66	2.000000	0.000000
3	45.500000	31.819805	14725.500000	6.000000	2	2.000000	0.000000

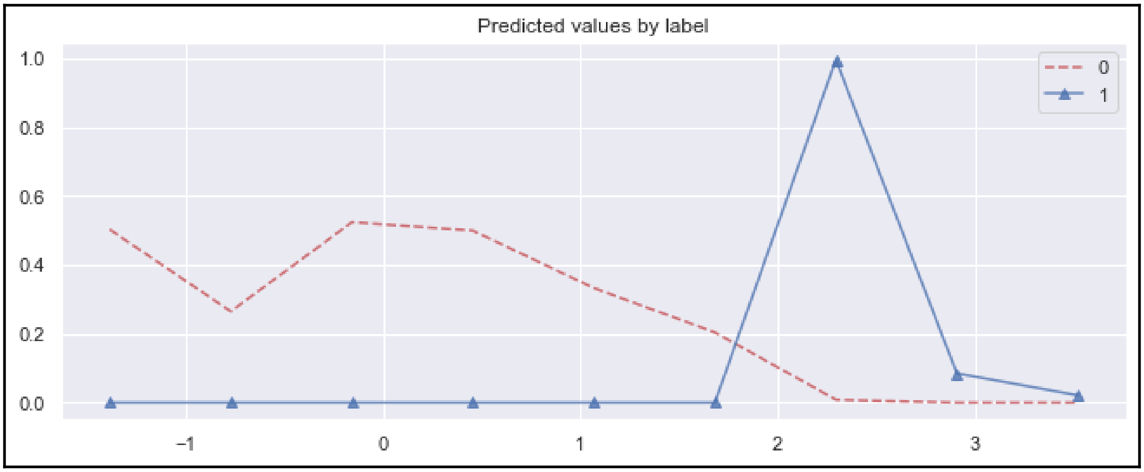
$$z = \frac{x - \mu}{\sigma}$$

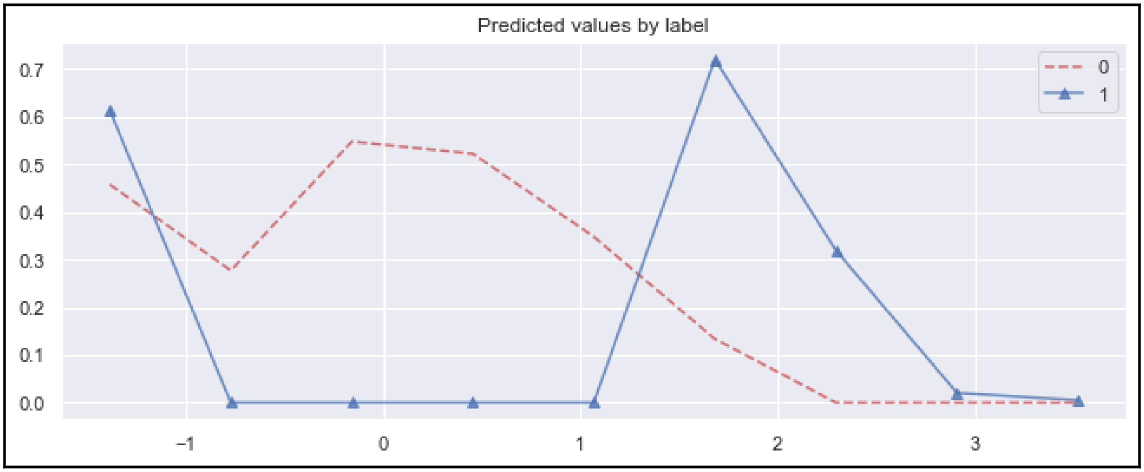


	timestamp	value	label
0	1469376000	0.847300	0
1	1469376300	-0.036137	0
2	1469376600	0.074292	0
3	1469376900	0.074292	0
4	1469377200	-0.036137	0




Values by label
1.0
0.8
0.6
0.4
0.2
0.0
−1
0
1
2
3
4
0
1


Predicted values by label
1.0
0.8
0.6
0.4
0.2
0.0
−1
0
1
2
3
0
1



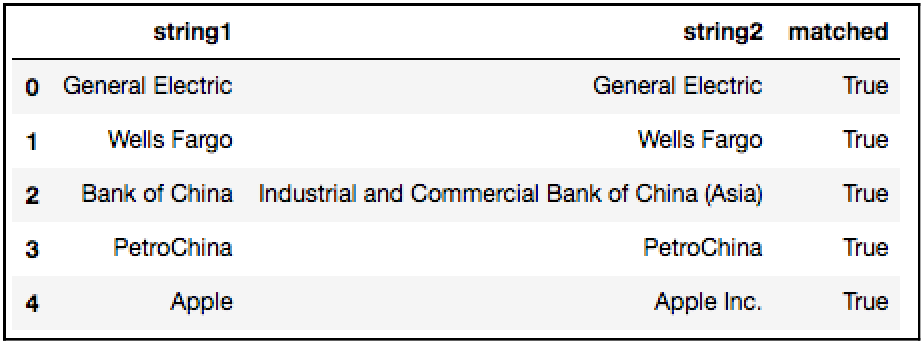
$$\phi : \mathcal{X} \rightarrow \mathcal{F}$$

$$\psi : \mathcal{F} \rightarrow \mathcal{X}$$

$$\phi$$

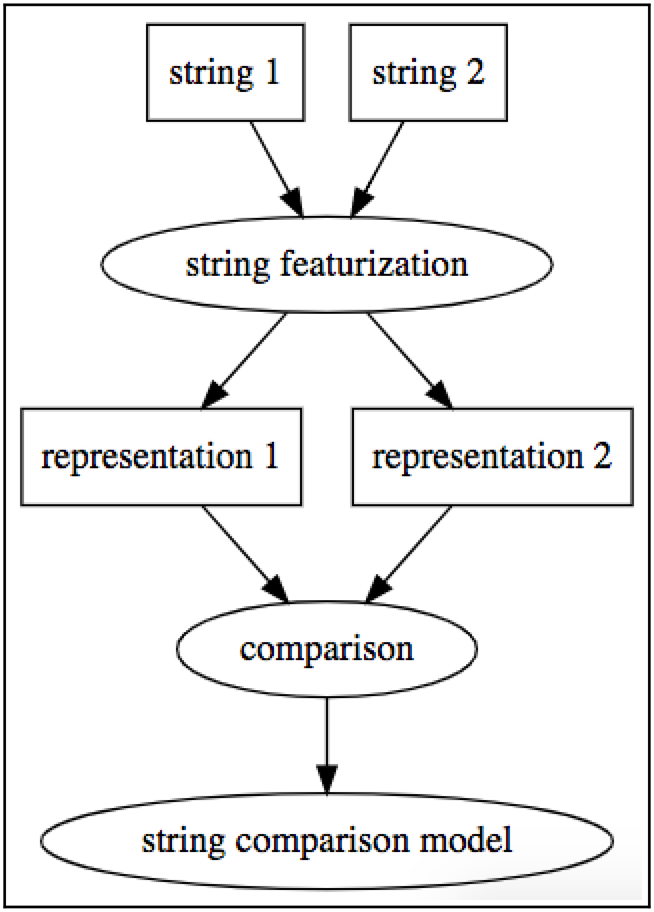
$$\psi$$

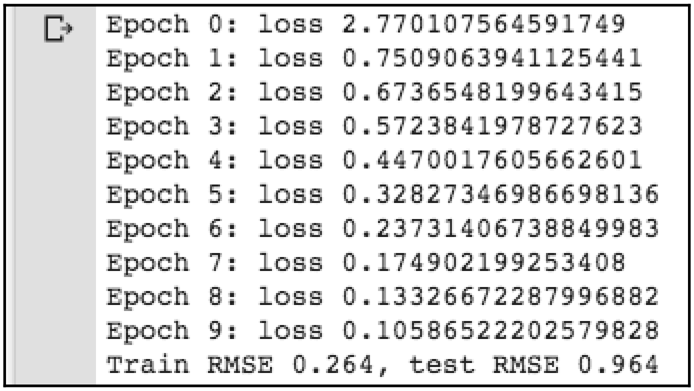
$$\phi, \psi = \underset{\phi, \psi}{\arg\min} \left\| X - (\psi \circ \phi) X \right\|^2$$



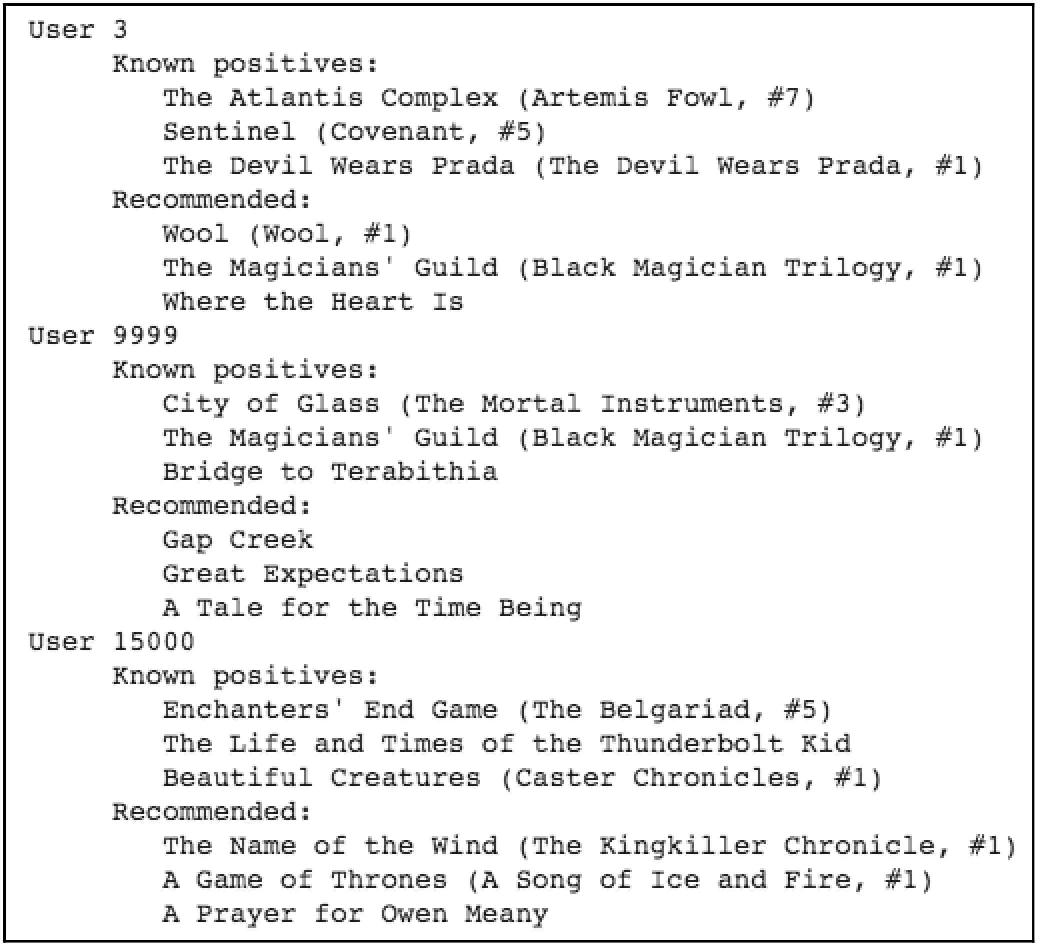
	string1	string2	matched
0	General Electric	General Electric	True
1	Wells Fargo	Wells Fargo	True
2	Bank of China	Industrial and Commercial Bank of China (Asia)	True
3	PetroChina	PetroChina	True
4	Apple	Apple Inc.	True

$$\sqrt[2]{\sum_{i} x_i^2}$$

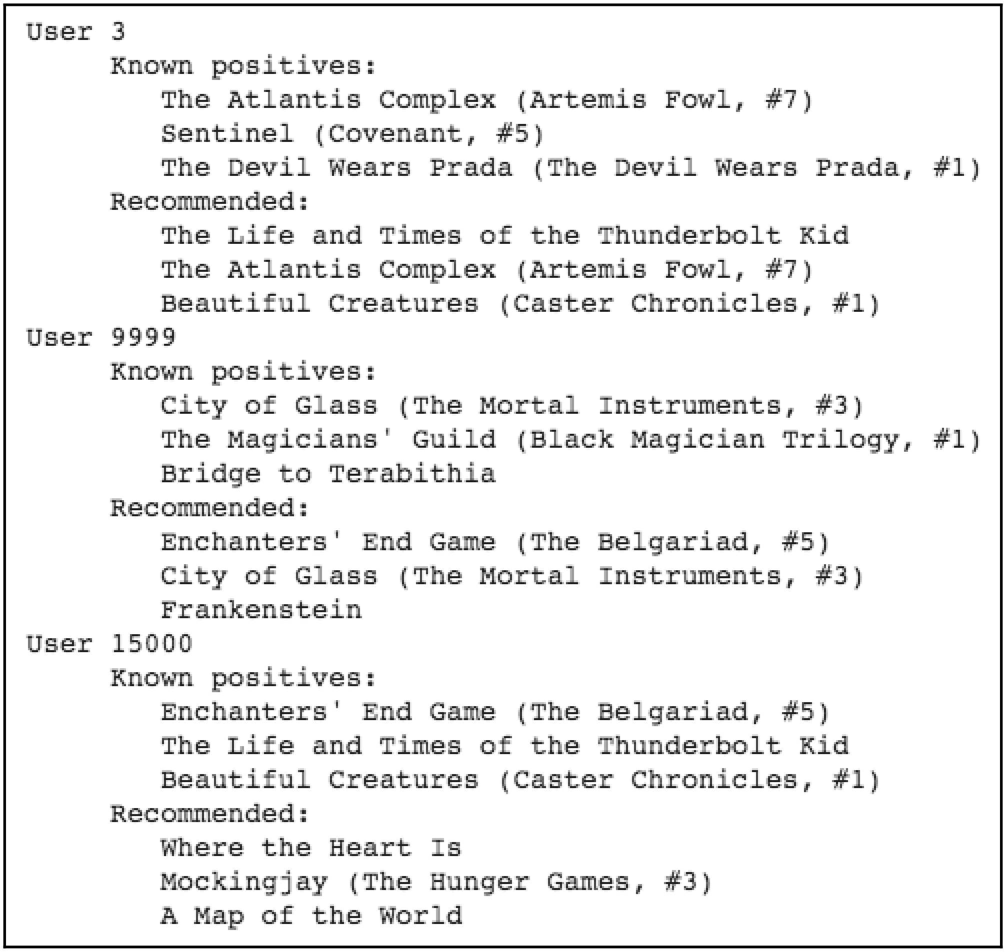




```
Epoch 0: loss 2.770107564591749
Epoch 1: loss 0.7509063941125441
Epoch 2: loss 0.6736548199643415
Epoch 3: loss 0.5723841978727623
Epoch 4: loss 0.4470017605662601
Epoch 5: loss 0.32827346986698136
Epoch 6: loss 0.23731406738849983
Epoch 7: loss 0.174902199253408
Epoch 8: loss 0.13326672287996882
Epoch 9: loss 0.10586522202579828
Train RMSE 0.264, test RMSE 0.964
```



```
User 3
     Known positives:
        The Atlantis Complex (Artemis Fowl, #7)
        Sentinel (Covenant, #5)
        The Devil Wears Prada (The Devil Wears Prada, #1)
     Recommended:
        Wool (Wool, #1)
        The Magicians' Guild (Black Magician Trilogy, #1)
        Where the Heart Is
User 9999
     Known positives:
        City of Glass (The Mortal Instruments, #3)
        The Magicians' Guild (Black Magician Trilogy, #1)
        Bridge to Terabithia
     Recommended:
        Gap Creek
        Great Expectations
        A Tale for the Time Being
User 15000
     Known positives:
        Enchanters' End Game (The Belgariad, #5)
        The Life and Times of the Thunderbolt Kid
        Beautiful Creatures (Caster Chronicles, #1)
     Recommended:
        The Name of the Wind (The Kingkiller Chronicle, #1)
        A Game of Thrones (A Song of Ice and Fire, #1)
        A Prayer for Owen Meany
```



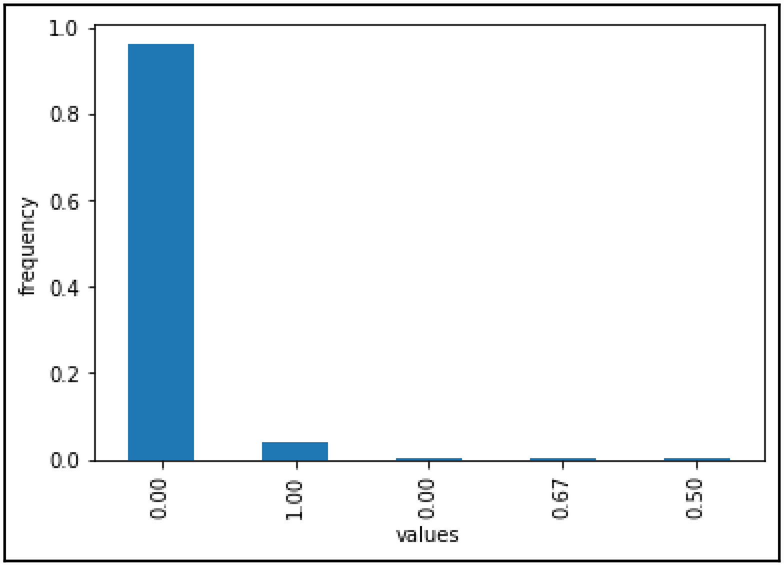
```
User 3
     Known positives:
        The Atlantis Complex (Artemis Fowl, #7)
        Sentinel (Covenant, #5)
        The Devil Wears Prada (The Devil Wears Prada, #1)
     Recommended:
        The Life and Times of the Thunderbolt Kid
        The Atlantis Complex (Artemis Fowl, #7)
        Beautiful Creatures (Caster Chronicles, #1)
User 9999
     Known positives:
        City of Glass (The Mortal Instruments, #3)
        The Magicians' Guild (Black Magician Trilogy, #1)
        Bridge to Terabithia
     Recommended:
        Enchanters' End Game (The Belgariad, #5)
        City of Glass (The Mortal Instruments, #3)
        Frankenstein
User 15000
     Known positives:
        Enchanters' End Game (The Belgariad, #5)
        The Life and Times of the Thunderbolt Kid
        Beautiful Creatures (Caster Chronicles, #1)
     Recommended:
        Where the Heart Is
        Mockingjay (The Hunger Games, #3)
        A Map of the World
```

$$\tilde{r}_{ui} = \sum_{f=0}^{nfactors} H_{u,f} W_{f,i}$$

$$\tilde{r}_{ui} = f(H_{u,f}W_{f,i}b_ib_u)$$

$$f$$

$$\prod_{(u,i)\in S_+} \tilde{r}_{ui} \prod_{(u,i)\in S_-} (1 - \tilde{r}_{ui})$$



$$O(n^2)$$

$$Q$$

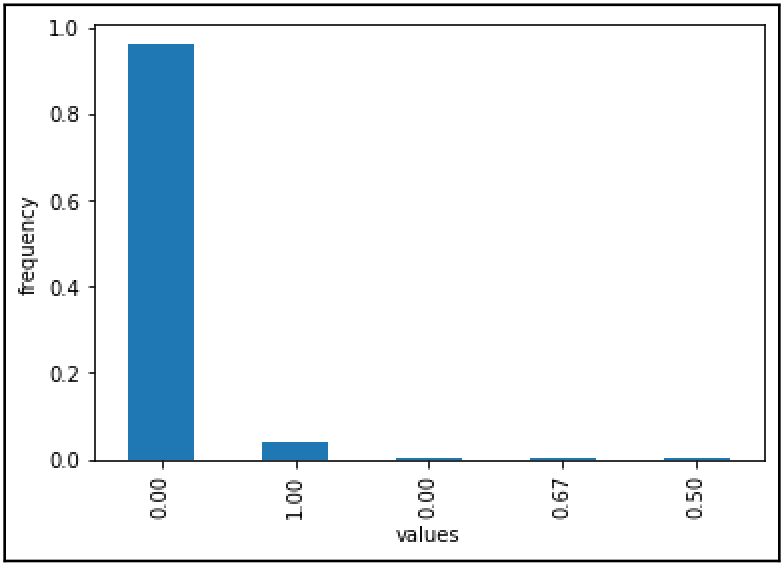
$$i$$

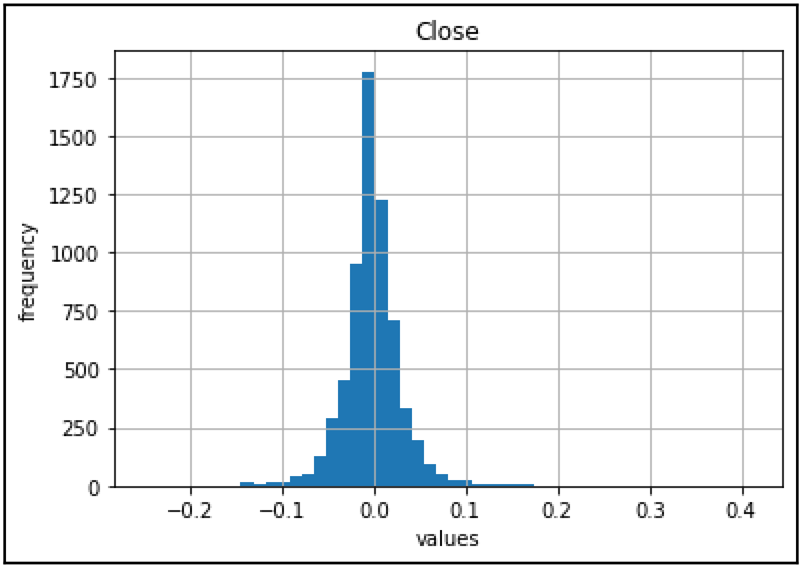
$$\Delta Q$$

$$O(m^2n)$$

$$X$$

$$\mathrm{H}(X) = -\sum_{i=1}^{n} \mathrm{P}(x_i) \log \mathrm{P}(x_i)$$


1.0
0.8
0.6
0.4
0.2
0.0
frequency
0.00
1.00
0.00
0.67
0.50
values



$$P(y=1|x) = \frac{1}{1+\exp(Af(x)+B)}$$

$$\operatorname{argmin}_{A,B} - \sum_i y_i \log(p_i) + (1-y_i)\log(1-p_i),$$

$$y_i = m(f(x_i)) + \epsilon_i,$$

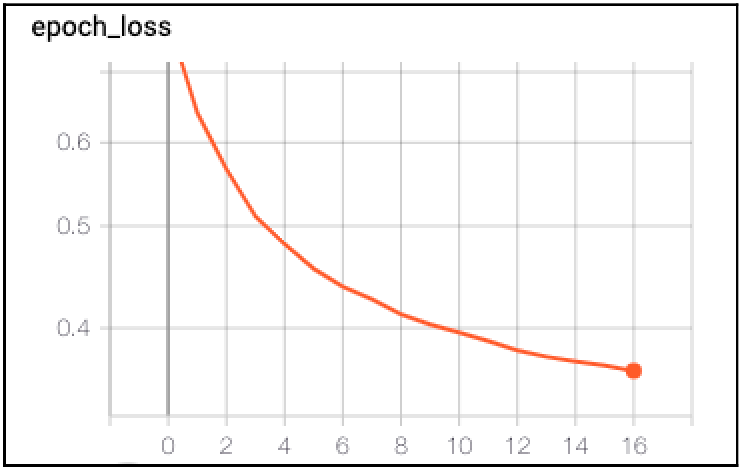
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

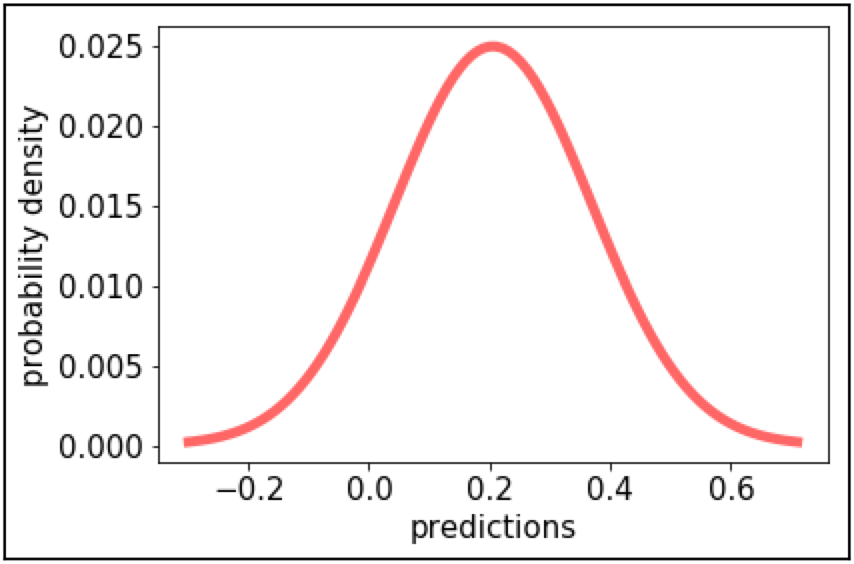
$$p(C_k \mid \mathbf{x}) = \frac{p(C_k)\ p(\mathbf{x} \mid C_k)}{p(\mathbf{x})}$$

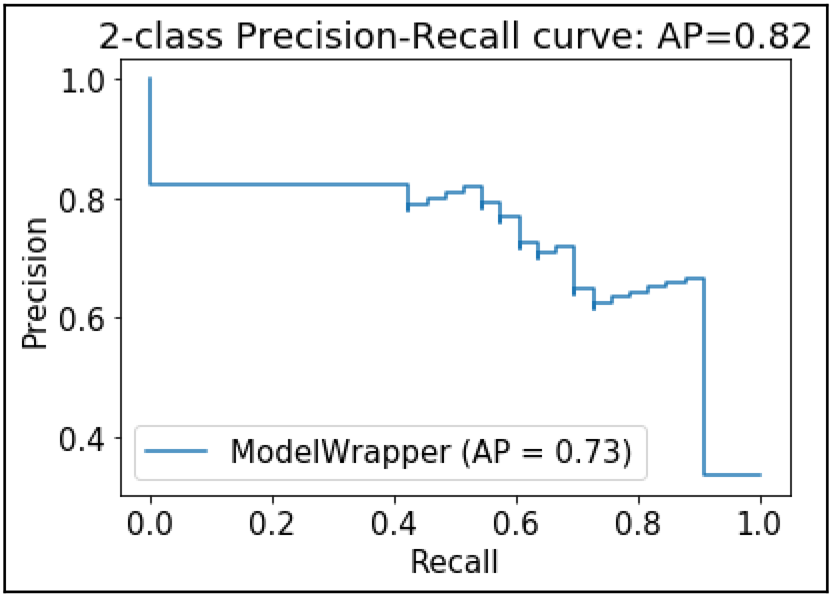
$$p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k)$$

$$\mathrm{CLV} = \mathrm{margin} \times \mathrm{transactions} \times \frac{revenue}{transaction}$$

$$m_x = \sum_{i=1}^{x} \frac{z_i}{x},$$

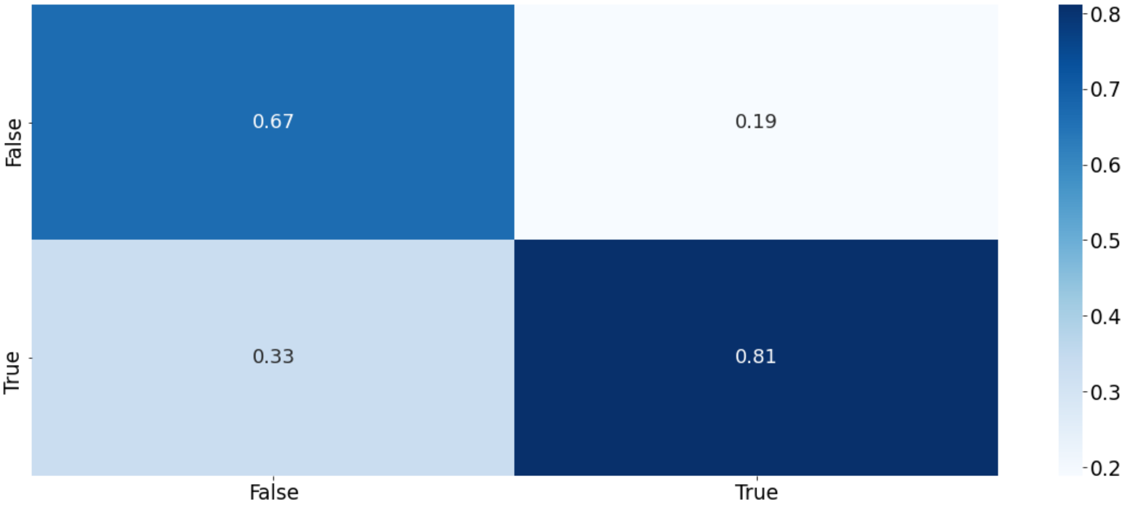


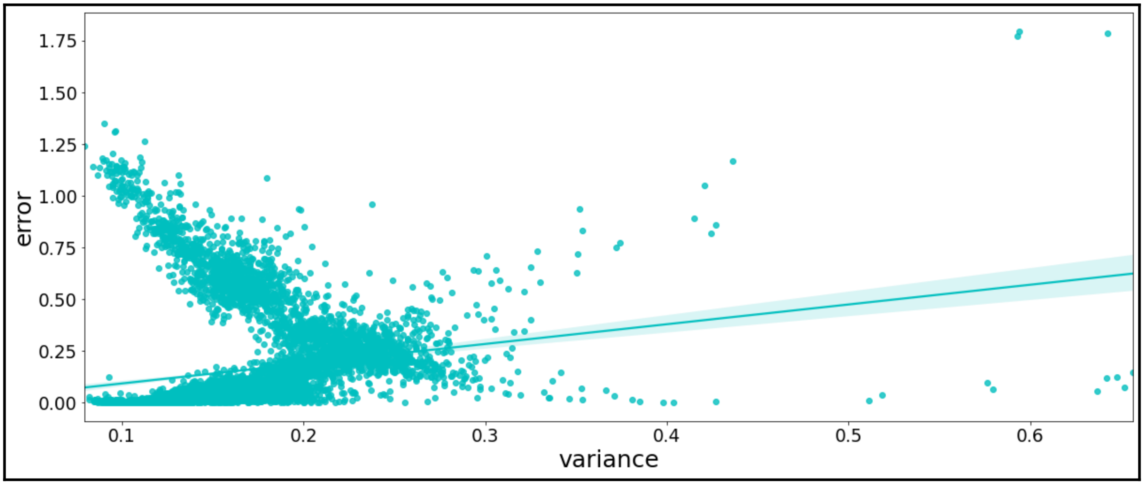

0.025
0.020
0.015
0.010
0.005
0.000
probability density
−0.2
0.0
0.2
0.4
0.6
predictions


2-class Precision-Recall curve: AP=0.82
1.0
0.8
0.6
0.4
Precision
ModelWrapper (AP = 0.73)
0.0
0.2
0.4
0.6
0.8
1.0
Recall

$$\text{Precision} = \frac{tp}{tp + fp}$$

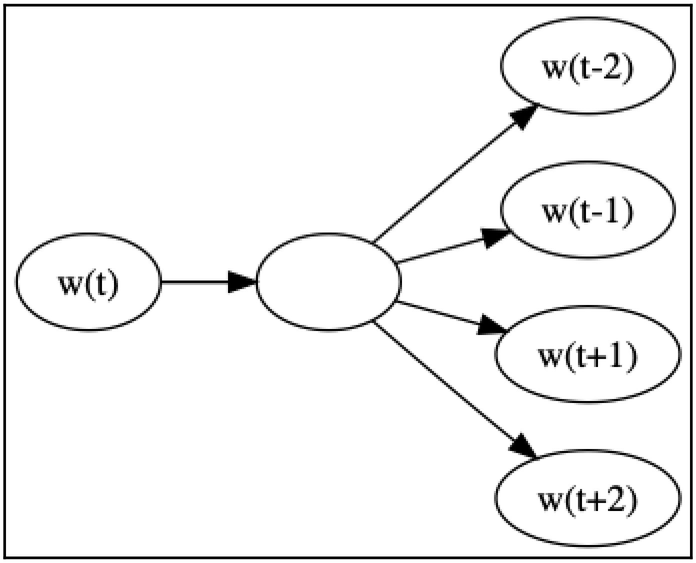
$$\text{Recall} = \frac{tp}{tp + fn}$$

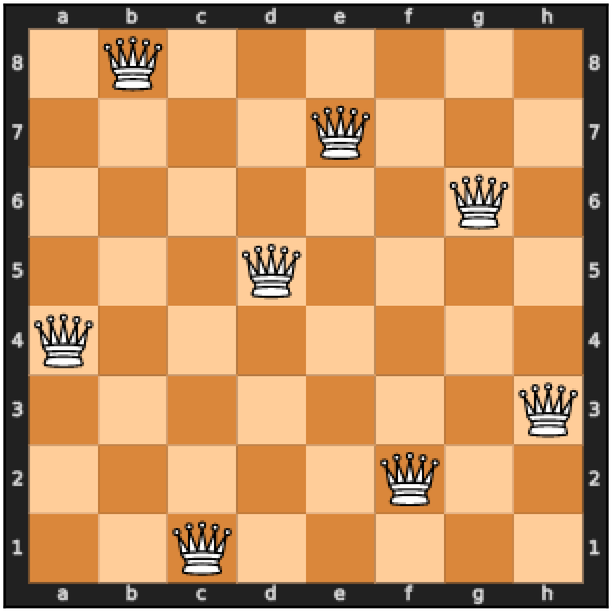


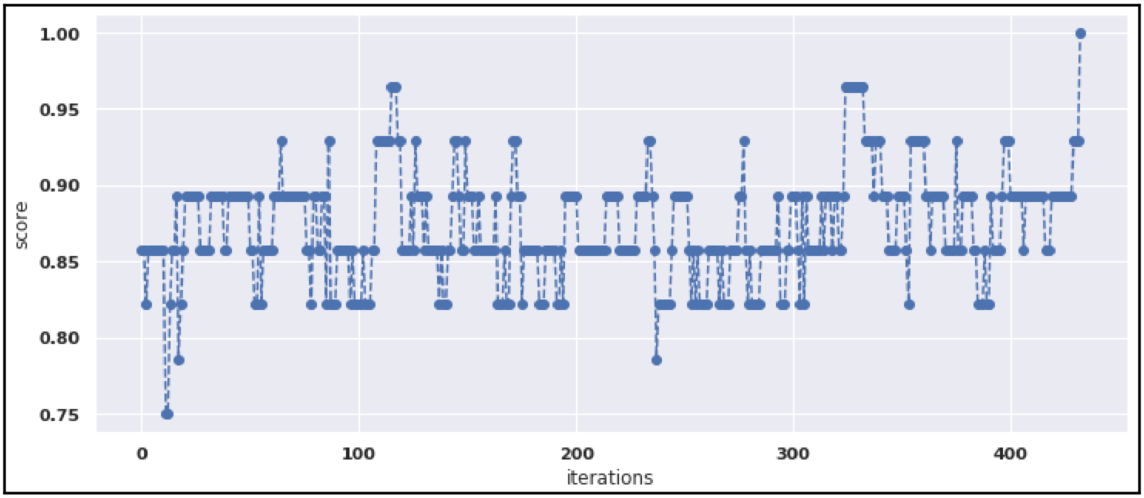

1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
error
0.1
0.2
0.3
0.4
0.5
0.6
variance

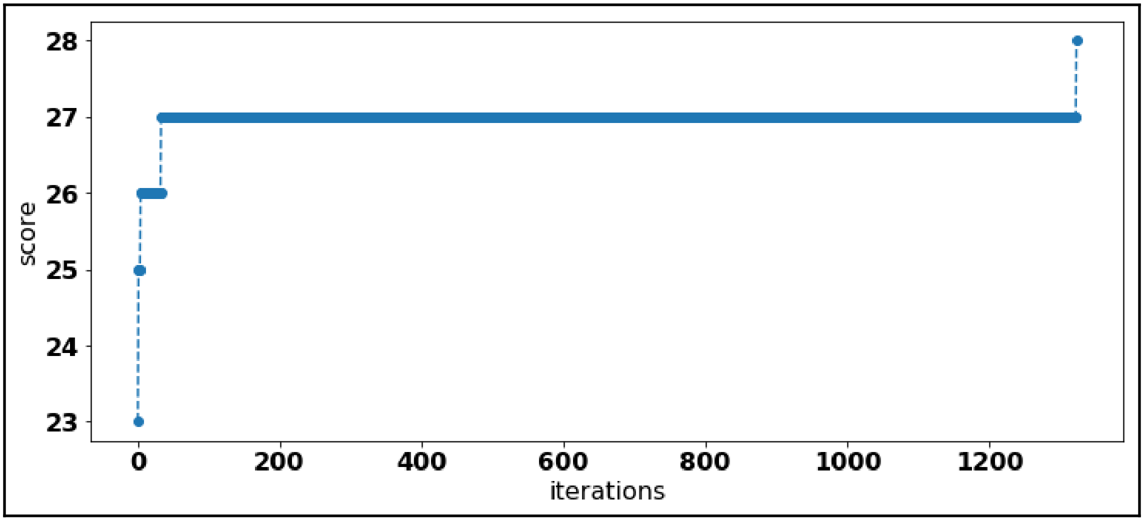
Chapter 5: Heuristic Search Techniques and Logical Inference

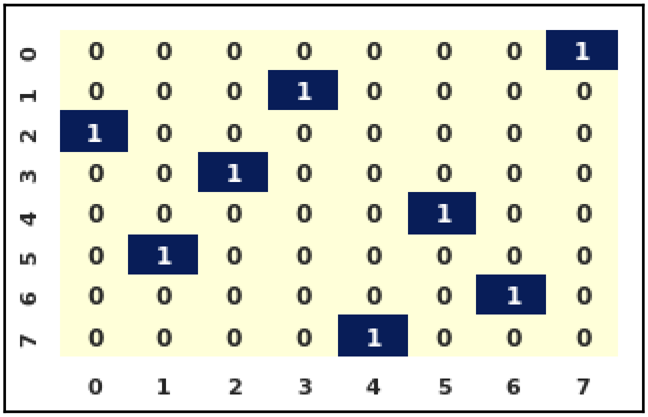
$$\begin{array}{ll} 1. & Ps \\ 2. & \forall x[Px \rightarrow Qx] \\ \hline \therefore & Qx \end{array}$$




a
b
c
d
e
f
g
h
8
7
6
5
4
3
2
1


1.00
0.95
0.90
0.85
0.80
0.75
score
0
100
200
300
400
iterations

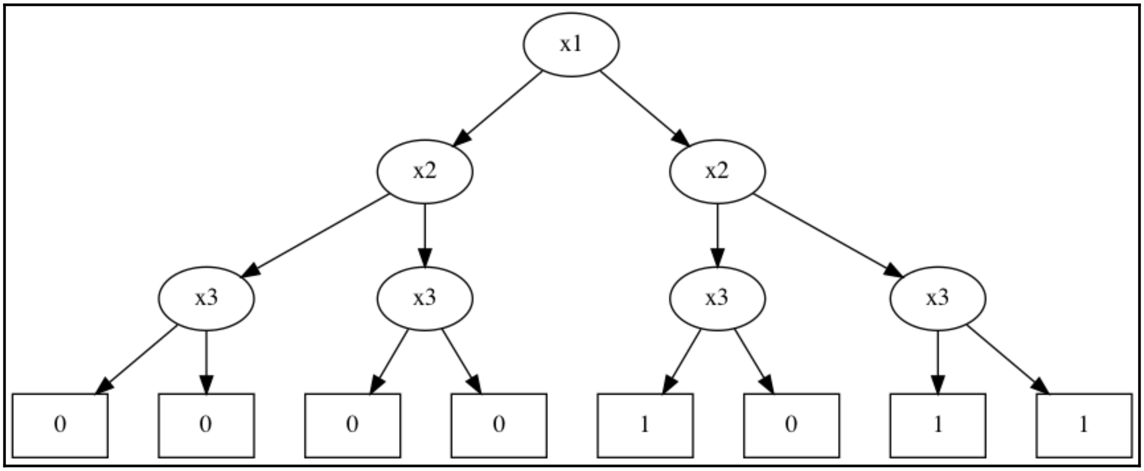


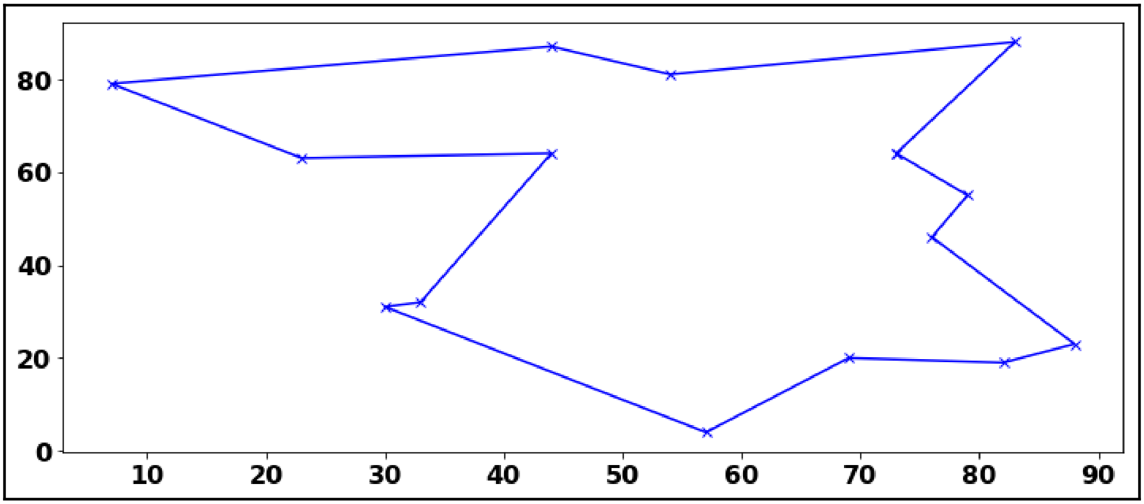


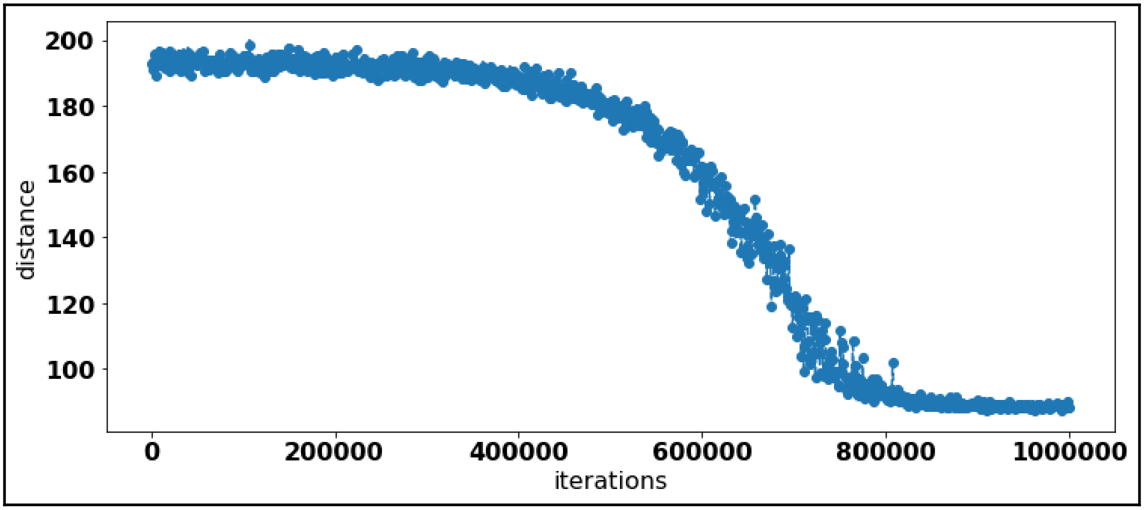
	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	1
1	0	0	0	1	0	0	0	0
2	1	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0
4	0	0	0	0	0	1	0	0
5	0	1	0	0	0	0	0	0
6	0	0	0	0	0	0	1	0
7	0	0	0	0	1	0	0	0

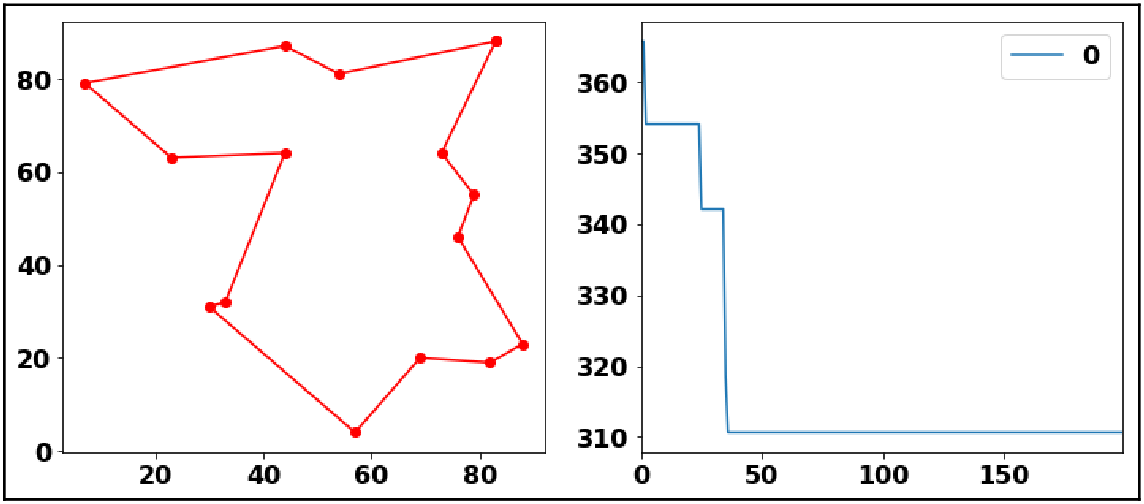
$$\sum_{i}^{N-1} i$$

$$f : \{0, 1\}^n \rightarrow \{0, 1\}$$


x1
x2
x2
x3
x3
x3
x3
0
0
0
0
1
0
1
1

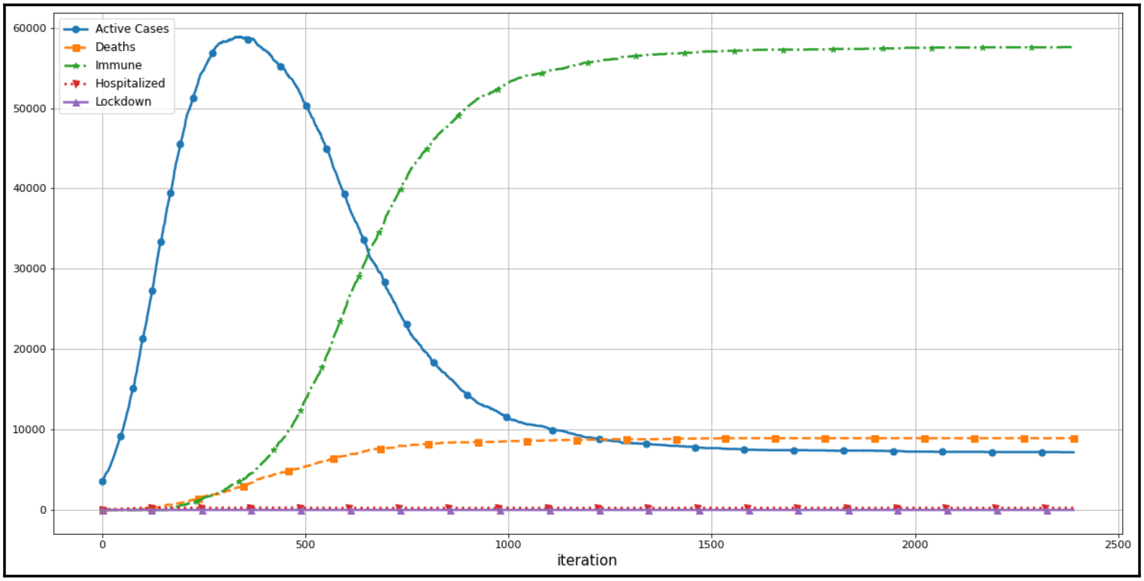

80
60
40
20
0
10
20
30
40
50
60
70
80
90

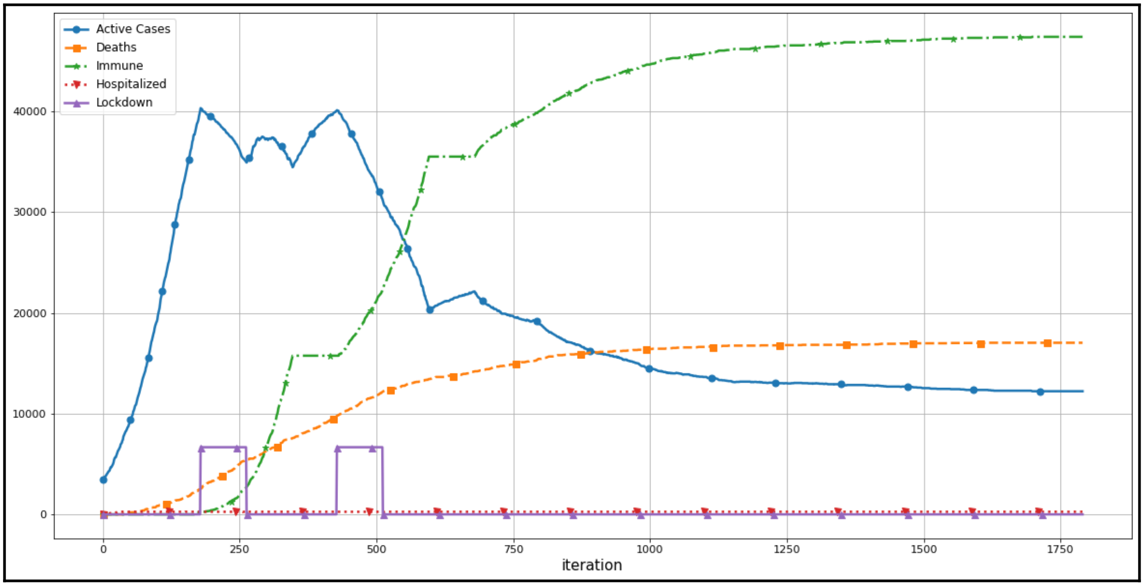


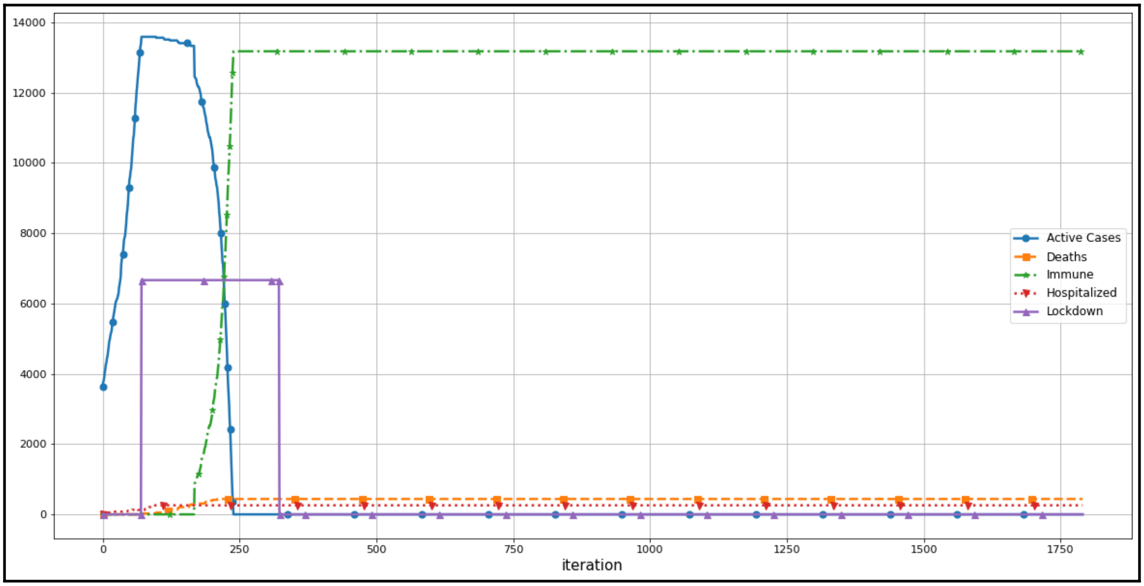


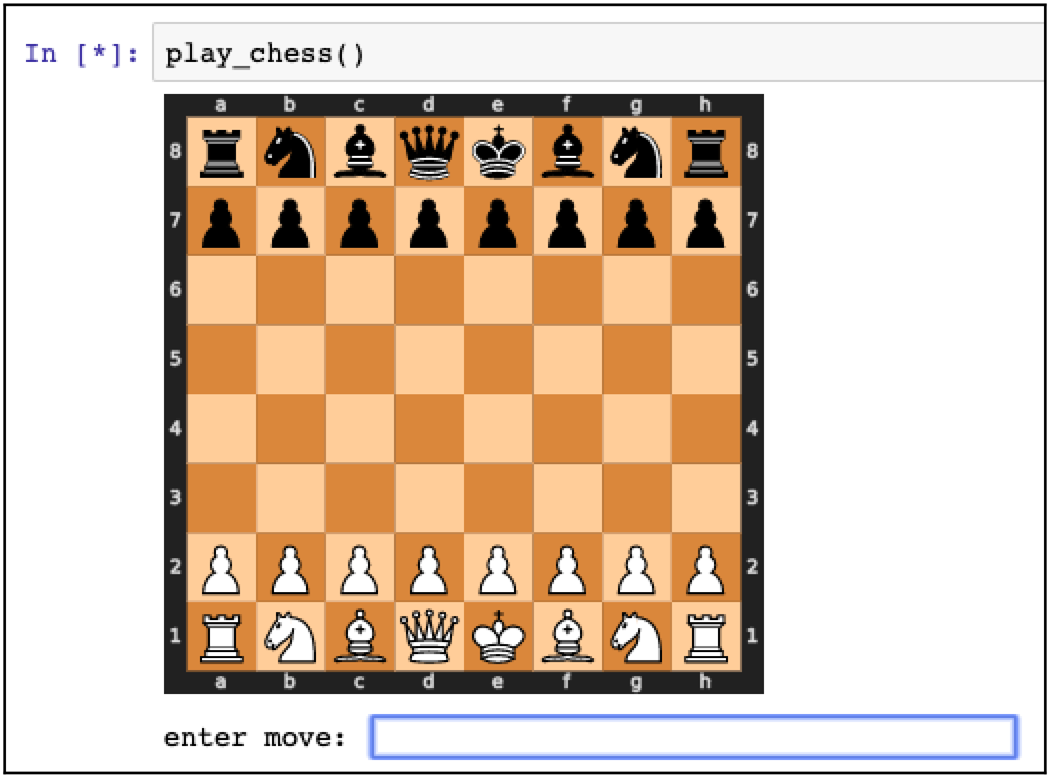
$$\exp((d_{\text{before}} - d_{\text{after}})/\text{tmp})$$

$$p_{xy}^{k} = \frac{(\tau_{xy}^{\alpha})(\eta_{xy}^{\beta})}{\sum_{z \in \mathrm{allowed}_x} (\tau_{xz}^{\alpha})(\eta_{xz}^{\beta})}$$




Active Cases
Deaths
Immune
Hospitalized
Lockdown
40000
30000
20000
10000
0
0
250
500
750
1000
1250
1500
1750
iteration

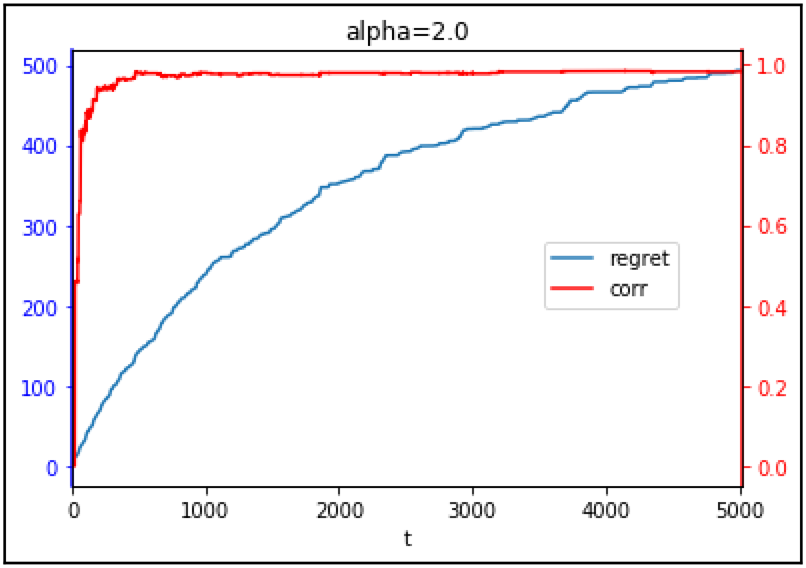

14000
12000
10000
8000
6000
4000
2000
0
Active Cases
Deaths
Immune
Hospitalized
Lockdown
0
250
500
750
1000
1250
1500
1750
iteration

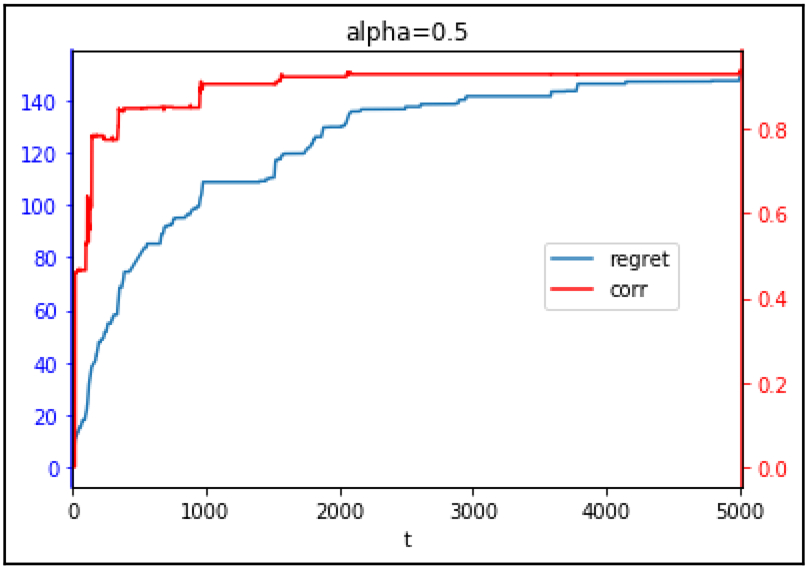


```
In [*]: play_chess()
```

```
enter move:
```

$$score(n) = \mathrm{wins}_n/\mathrm{nsims}_n + c \times \sqrt{\log(2 + \mathrm{nsims})/\mathrm{nsims}_n}$$


alpha=2.0
500
400
300
200
100
0
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
t
regret
corr


alpha=0.5
140
120
100
80
60
40
20
0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
t
regret
corr

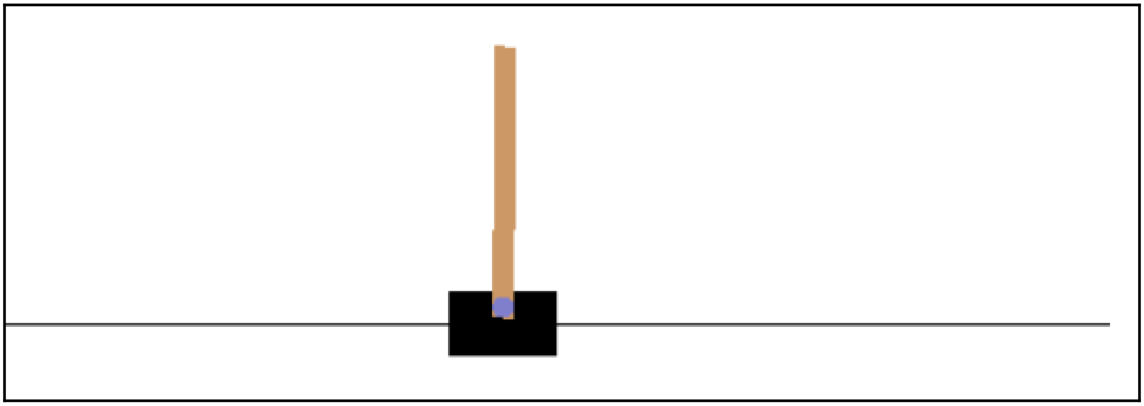
$$\alpha' \leftarrow \mathrm{argmax}_{a \in A} Q(a) + \sqrt{\frac{\alpha \ln t}{N(a)}},$$

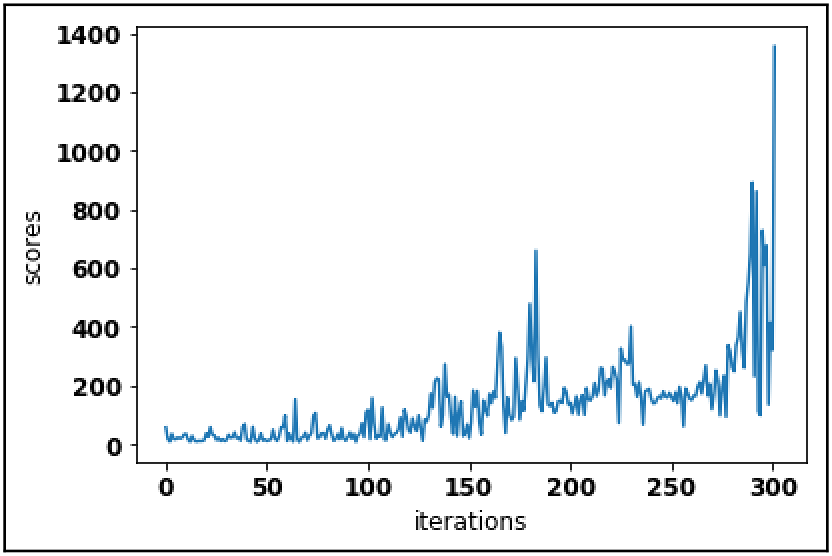
$$\boldsymbol{Q(a)}$$

$$\boldsymbol{N(a)}$$

$$a$$

$$\alpha > 0$$




1400
1200
1000
800
600
400
200
0
scores
0
50
100
150
200
250
300
iterations

$$s$$

$$a$$

$$\boldsymbol{N}$$

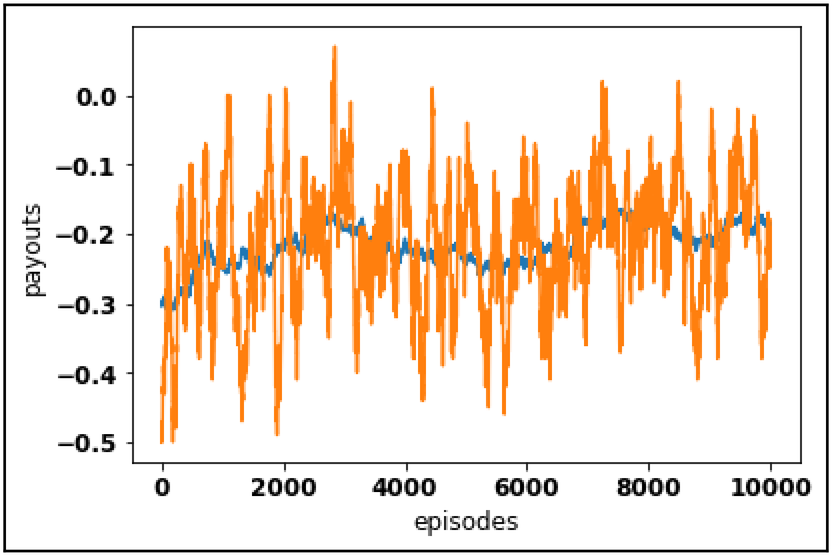
$$\pi(a_t = a_i, s_{t-1} = \frac{e^{p(a_i, s_{t-1})}}{\sum_{j=0}^{N} e^{p(a_j, s_{t-1})}}$$

$$Q(a_t, s_t) \in \mathbb{R}$$

$$\boldsymbol{v}$$

$$\nabla\theta_t = \alpha\nabla_\theta \log \pi(a_t, s_t)Q(a, s)$$

$$\alpha$$



$$Q^{\pi}(s,a) : S \times A \rightarrow \mathbb{R}$$

$$Q^{\pi}(s,a) = \mathbb{E}[\sum_{k=0}^{\infty} \gamma^{k} r_{t+1|s_t=s,a_t=a,\pi}]$$

$$r_t, s_t \in S, a_t \in A$$

$$t$$

$$\gamma$$

$$\pi$$

$$Q$$

$$Q^*(s,a) = \max_{\pi \in \Pi} Q^{\pi}(s,a)$$

$$\pi^*(s) = \text{argmax}_{a \in A} Q^*(s,a).$$

$$J = (Q(s,a;\theta) - Y_k^Q)^2, \text{where}$$
$$Y_k^Q = r + \gamma \max_{a' \in A} Q(s', a'; \theta_k),$$

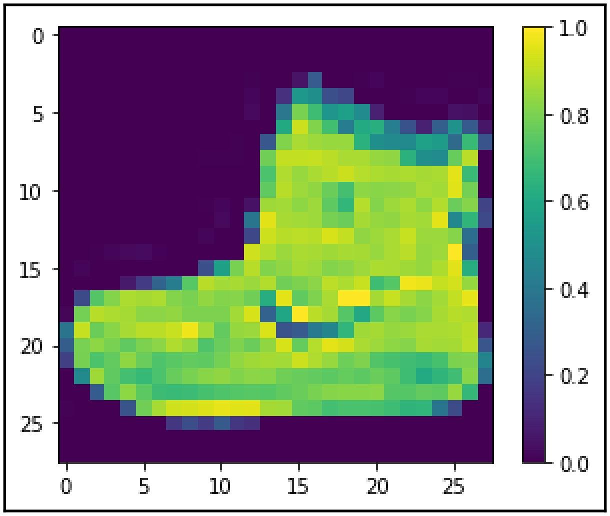
$$\boxed{\begin{aligned} J &= (Q(s,a;\theta) - Y_k^Q)^2, \text{where} \\ Y_k^Q &= r + \gamma \max_{a' \in A} Q(s', a'; \theta_k), \end{aligned}}$$

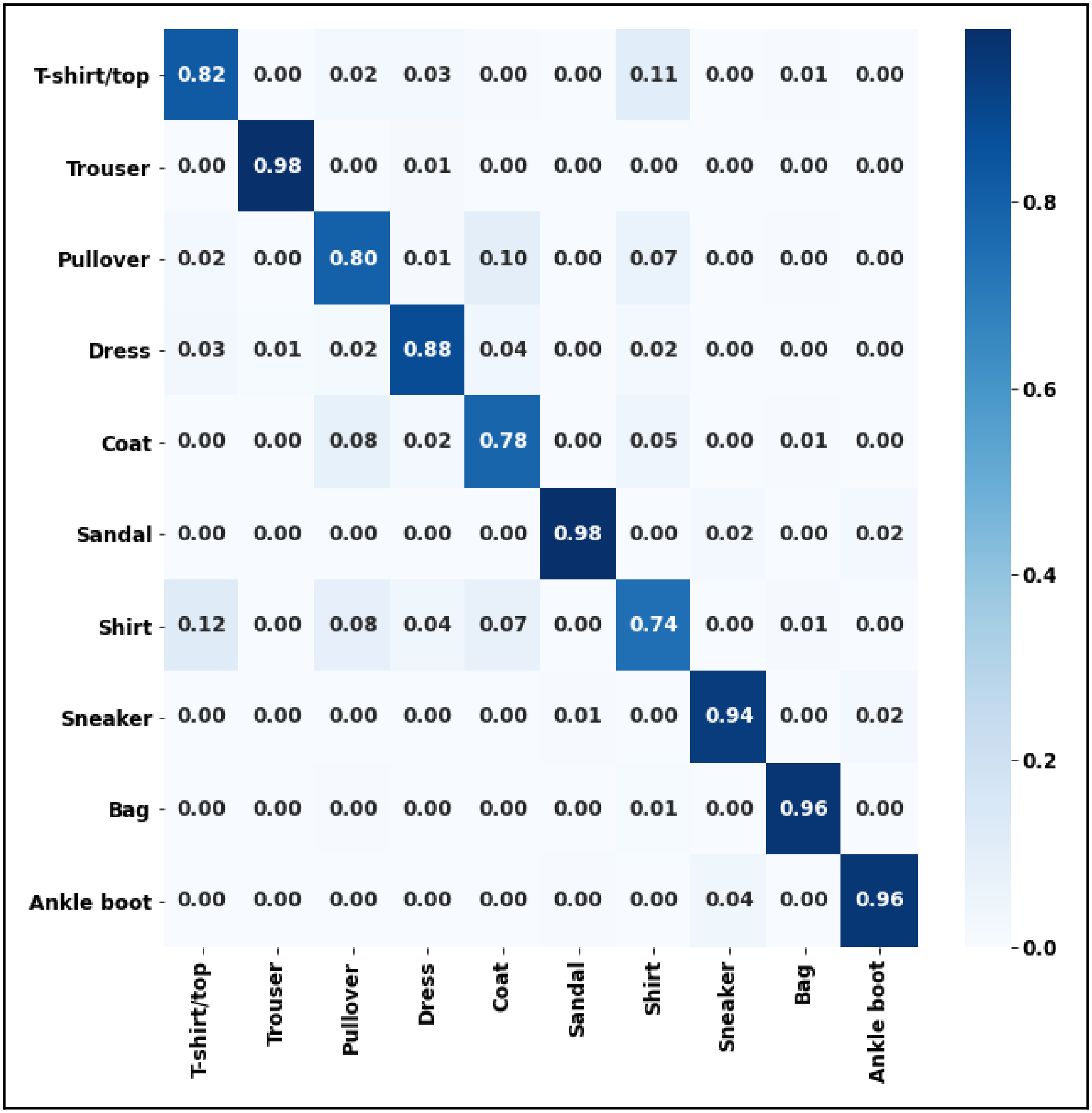
$$\theta_k$$

$$\boldsymbol{k}$$

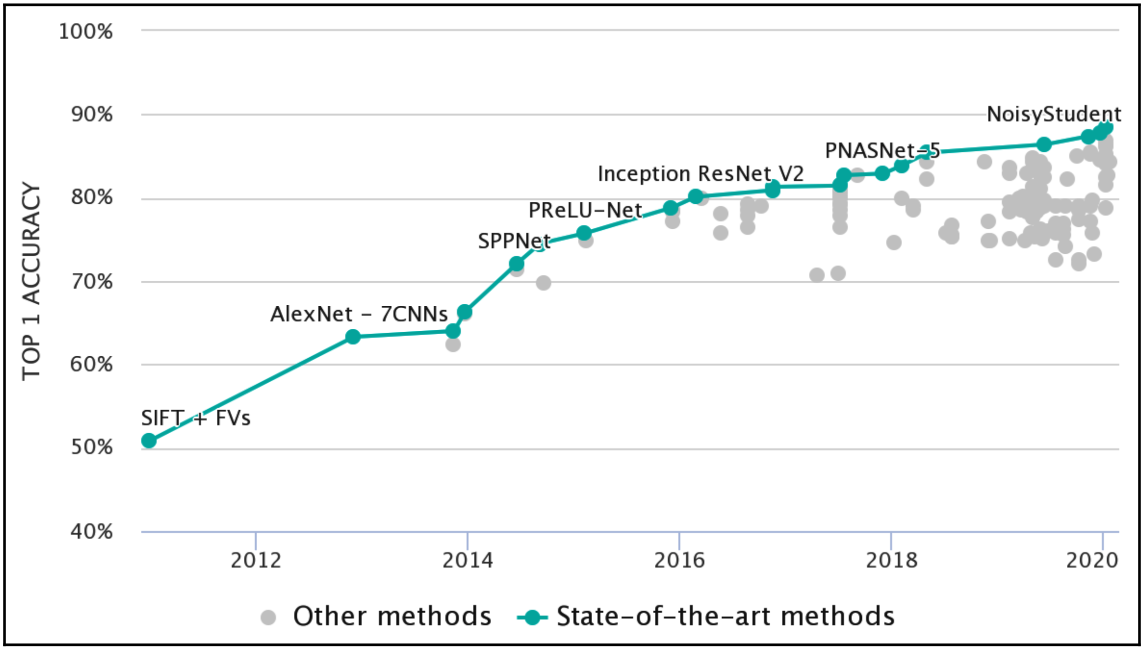
$$a', s'$$

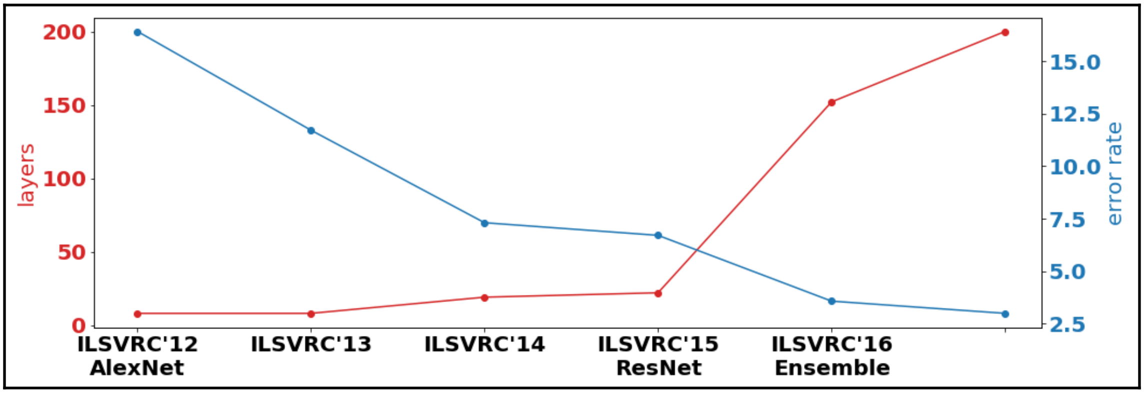

Ankle boot
T-shirt/top
T-shirt/top
Dress
T-shirt/top
Pullover
Sneaker
Pullover
Sandal
Sandal
T-shirt/top
Ankle boot
Sandal
Sandal
Sneaker
Ankle boot
Trouser
T-shirt/top
Shirt
Coat
Dress
Trouser
Coat
Bag
Coat

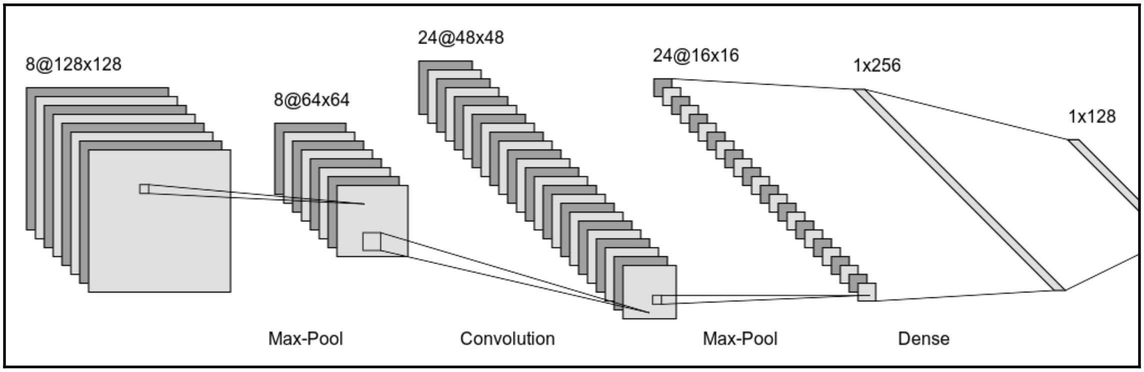

0
5
10
15
20
25
0
5
10
15
20
25
1.0
0.8
0.6
0.4
0.2
0.0

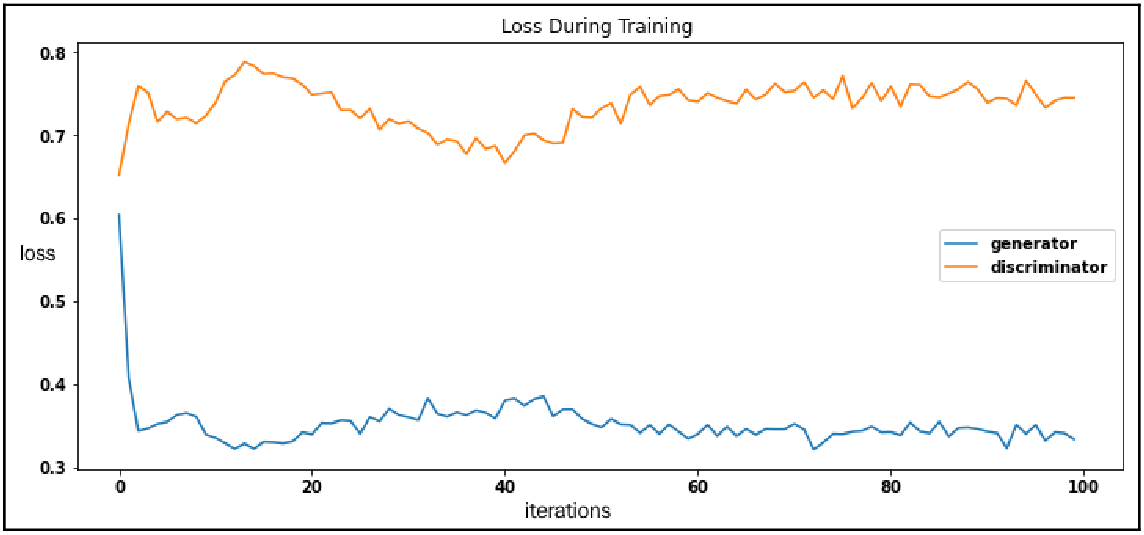


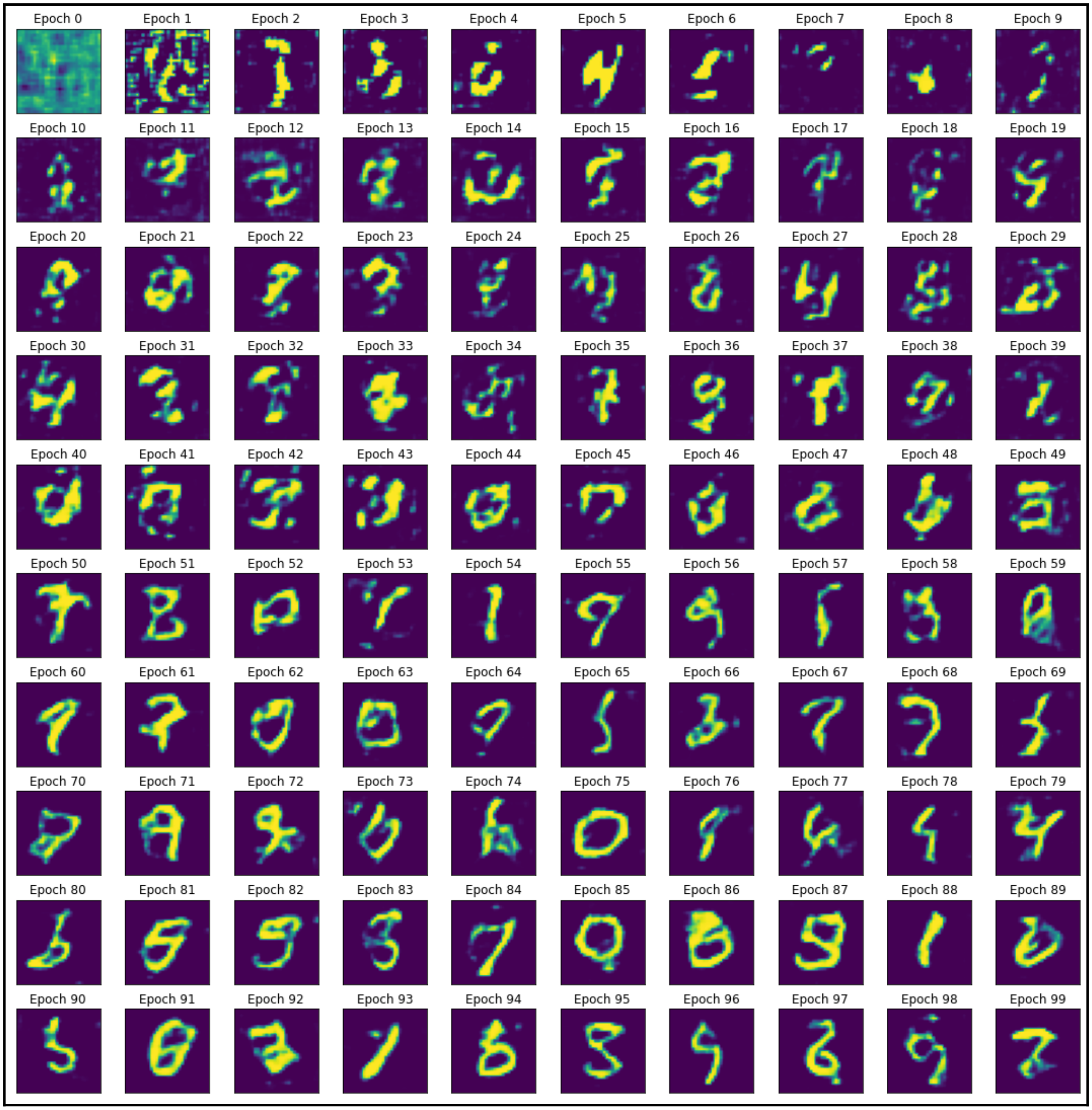
	T-shirt/top	Trouser	Pullover	Dress	Coat	Sandal	Shirt	Sneaker	Bag	Ankle boot
T-shirt/top	0.82	0.00	0.02	0.03	0.00	0.00	0.11	0.00	0.01	0.00
Trouser	0.00	0.98	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Pullover	0.02	0.00	0.80	0.01	0.10	0.00	0.07	0.00	0.00	0.00
Dress	0.03	0.01	0.02	0.88	0.04	0.00	0.02	0.00	0.00	0.00
Coat	0.00	0.00	0.08	0.02	0.78	0.00	0.05	0.00	0.01	0.00
Sandal	0.00	0.00	0.00	0.00	0.00	0.98	0.00	0.02	0.00	0.02
Shirt	0.12	0.00	0.08	0.04	0.07	0.00	0.74	0.00	0.01	0.00
Sneaker	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.94	0.00	0.02
Bag	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.96	0.00
Ankle boot	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.96

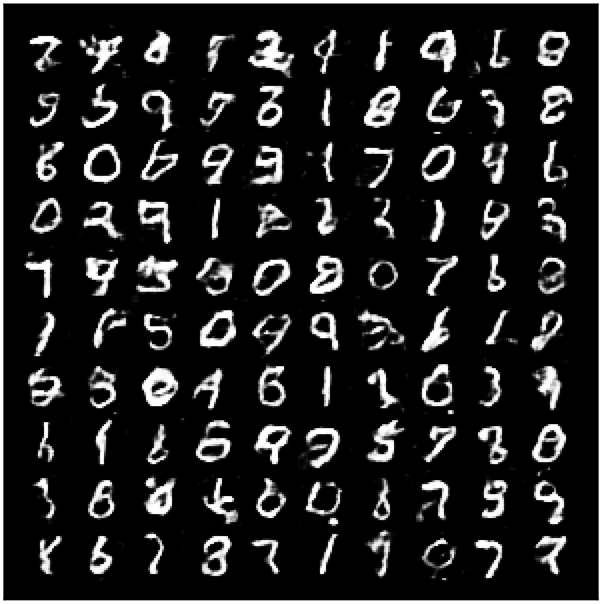

100%
90%
80%
70%
60%
50%
40%
TOP 1 ACCURACY
2012
2014
2016
2018
2020
SIFT + FVs
AlexNet – 7CNNs
SPPNet
PReLU-Net
Inception ResNet V2
PNASNet-5
NoisyStudent
● Other methods
State-of-the-art methods

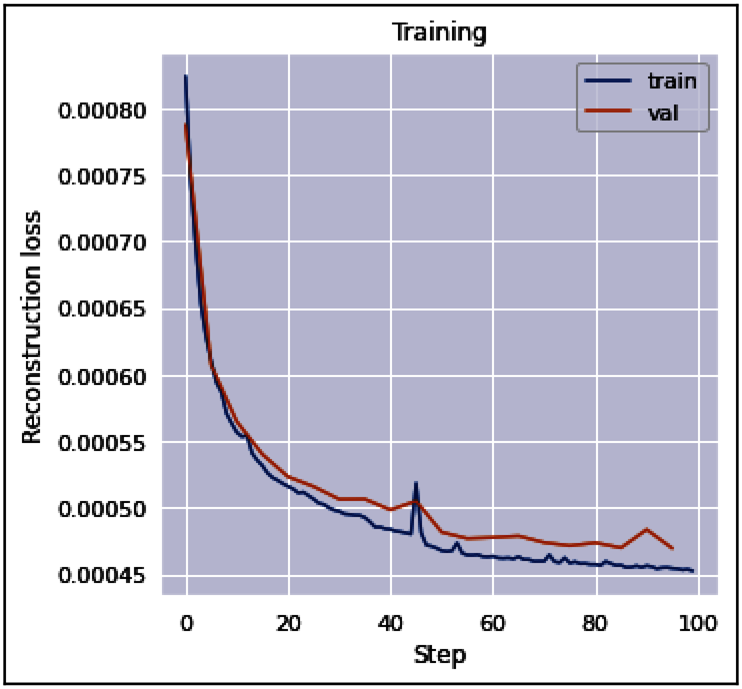

layers
200
150
100
50
0
error rate
15.0
12.5
10.0
7.5
5.0
2.5
ILSVRC'12
AlexNet
ILSVRC'13
ILSVRC'14
ILSVRC'15
ResNet
ILSVRC'16
Ensemble

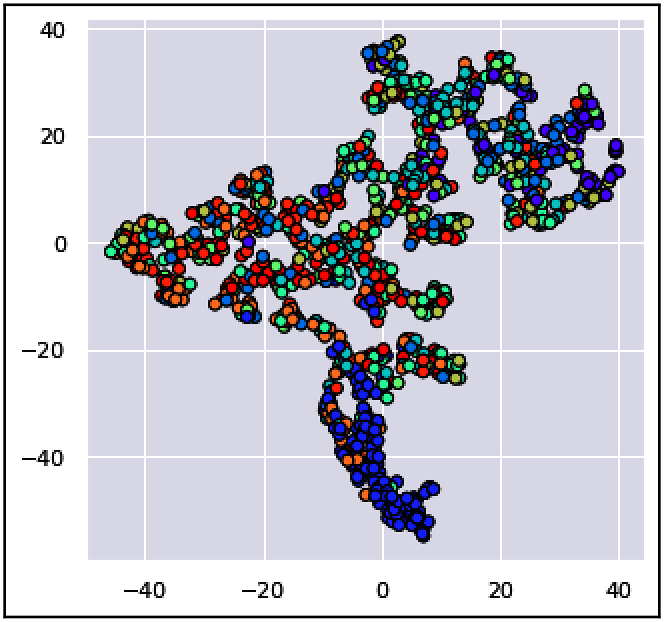

8@128x128
8@64x64
24@48x48
24@16x16
1x256
1x128
Max-Pool
Convolution
Max-Pool
Dense

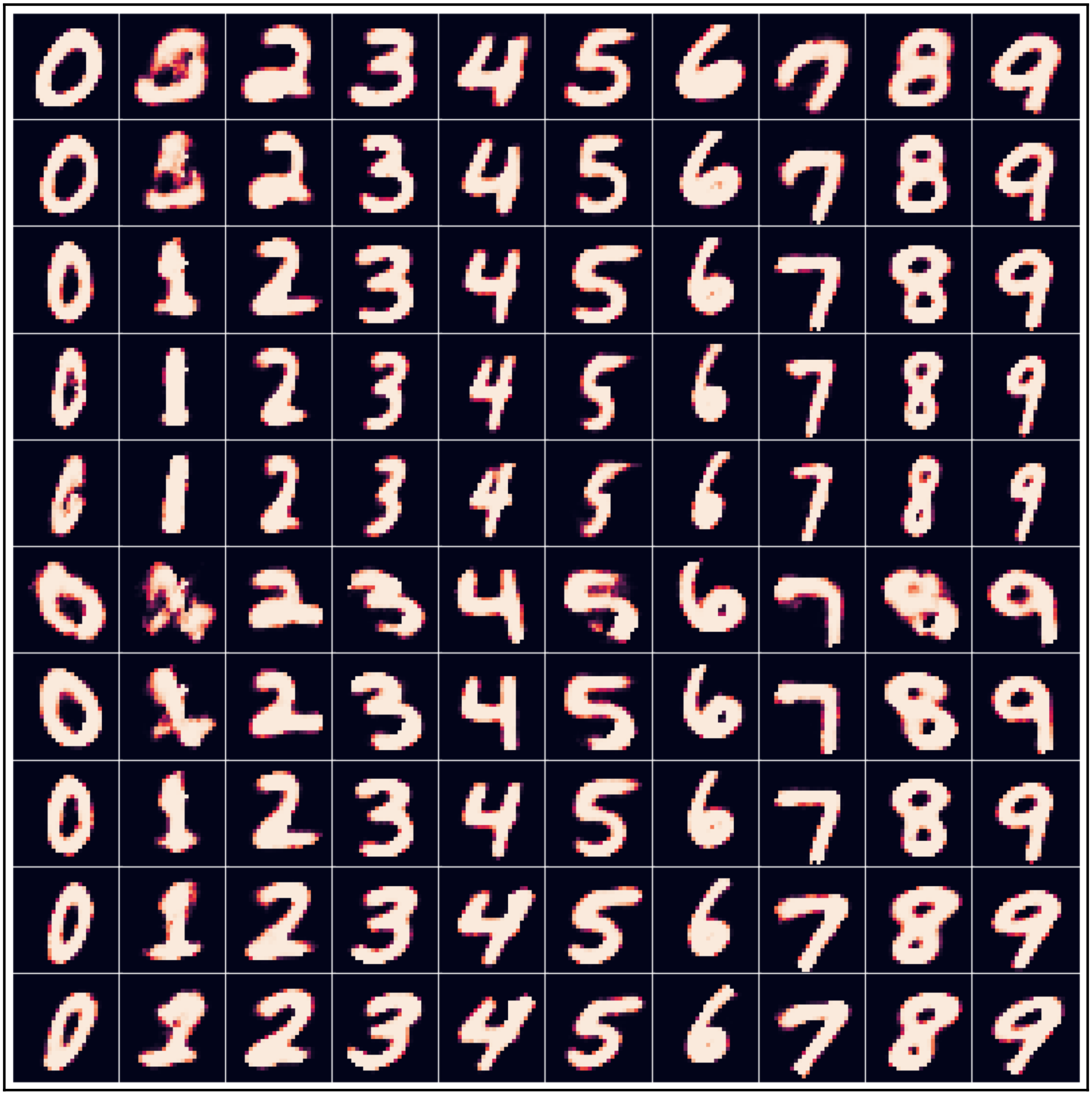

Loss During Training
loss
0.8
0.7
0.6
0.5
0.4
0.3
0
20
40
60
80
100
iterations
generator
discriminator

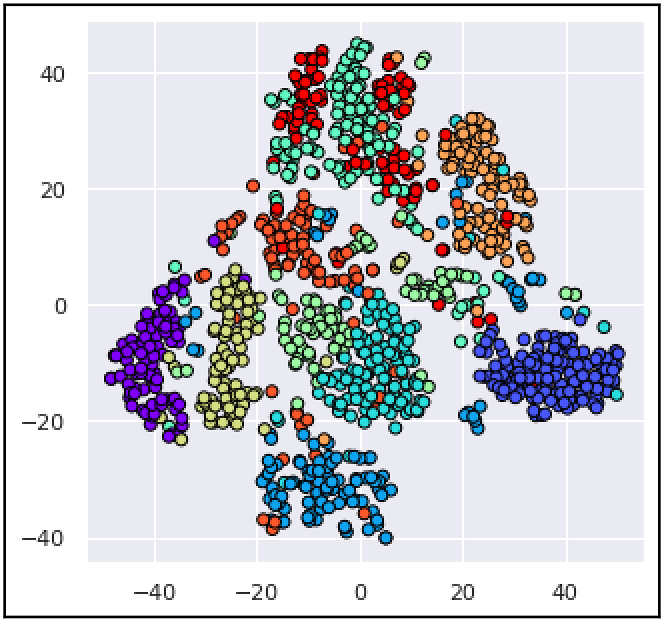

Epoch 0
Epoch 1
Epoch 2
Epoch 3
Epoch 4
Epoch 5
Epoch 6
Epoch 7
Epoch 8
Epoch 9
Epoch 10
Epoch 11
Epoch 12
Epoch 13
Epoch 14
Epoch 15
Epoch 16
Epoch 17
Epoch 18
Epoch 19
Epoch 20
Epoch 21
Epoch 22
Epoch 23
Epoch 24
Epoch 25
Epoch 26
Epoch 27
Epoch 28
Epoch 29
Epoch 30
Epoch 31
Epoch 32
Epoch 33
Epoch 34
Epoch 35
Epoch 36
Epoch 37
Epoch 38
Epoch 39
Epoch 40
Epoch 41
Epoch 42
Epoch 43
Epoch 44
Epoch 45
Epoch 46
Epoch 47
Epoch 48
Epoch 49
Epoch 50
Epoch 51
Epoch 52
Epoch 53
Epoch 54
Epoch 55
Epoch 56
Epoch 57
Epoch 58
Epoch 59
Epoch 60
Epoch 61
Epoch 62
Epoch 63
Epoch 64
Epoch 65
Epoch 66
Epoch 67
Epoch 68
Epoch 69
Epoch 70
Epoch 71
Epoch 72
Epoch 73
Epoch 74
Epoch 75
Epoch 76
Epoch 77
Epoch 78
Epoch 79
Epoch 80
Epoch 81
Epoch 82
Epoch 83
Epoch 84
Epoch 85
Epoch 86
Epoch 87
Epoch 88
Epoch 89
Epoch 90
Epoch 91
Epoch 92
Epoch 93
Epoch 94
Epoch 95
Epoch 96
Epoch 97
Epoch 98
Epoch 99

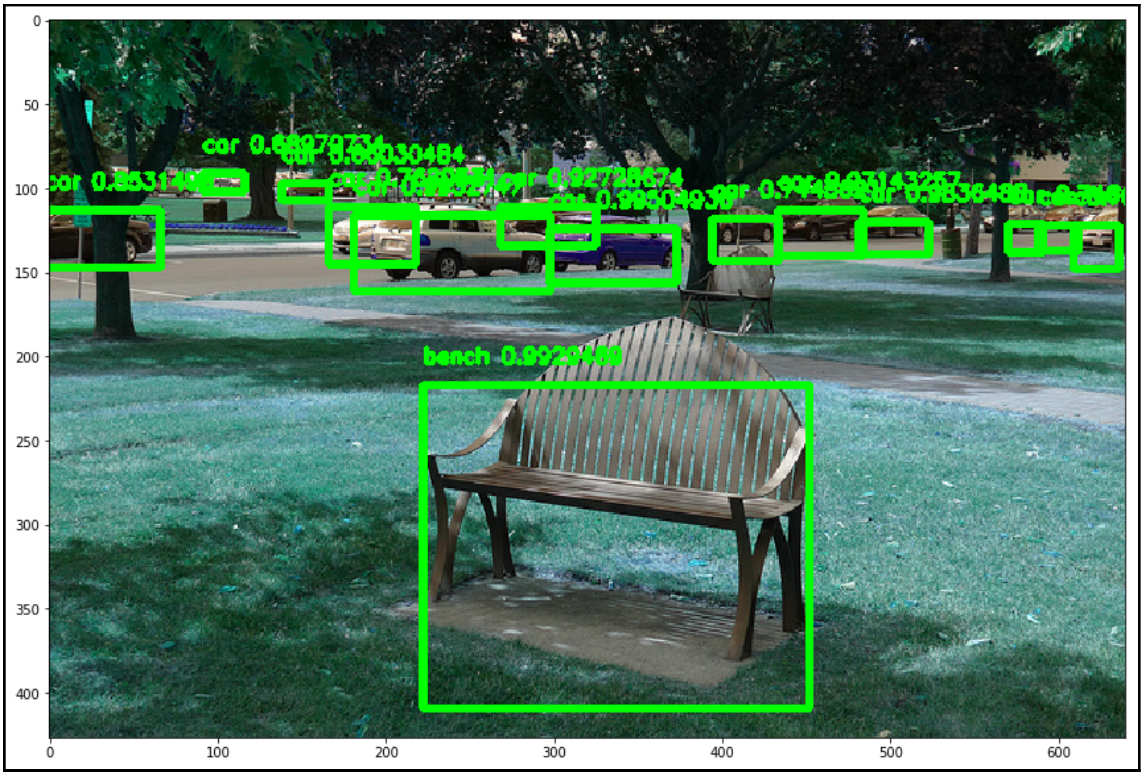


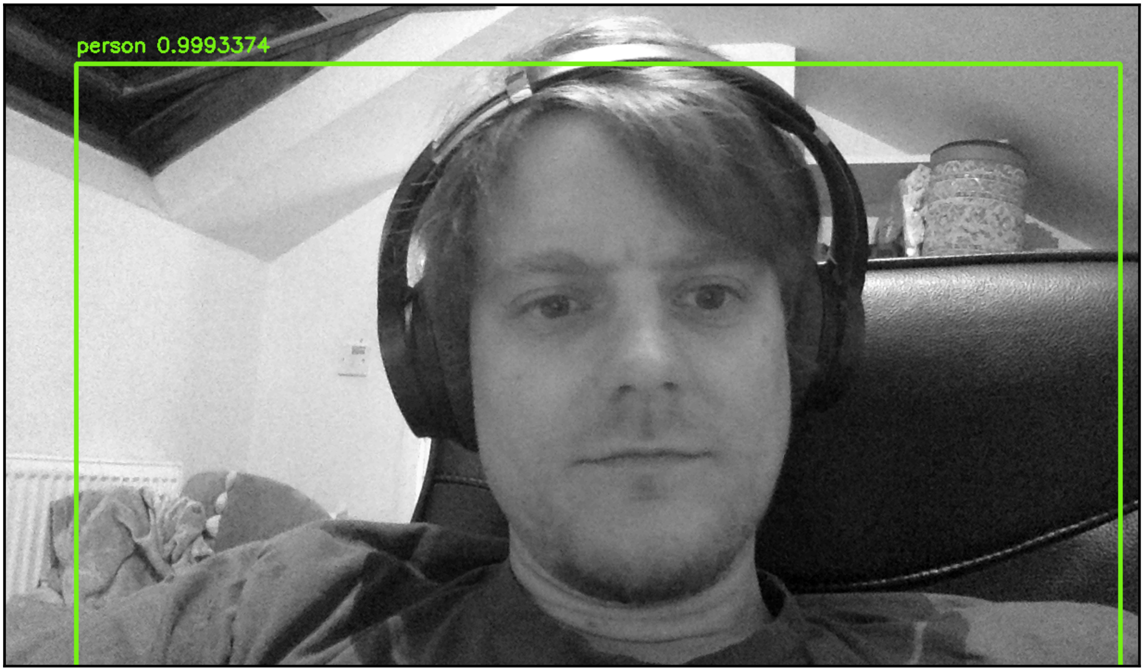

Training
train
val
0.00080
0.00075
0.00070
0.00065
0.00060
0.00055
0.00050
0.00045
Reconstruction loss
0
20
40
60
80
100
Step


40
20
0
−20
−40
−40
−20
0
20
40

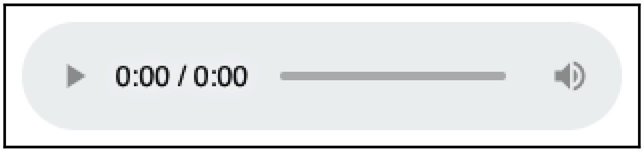


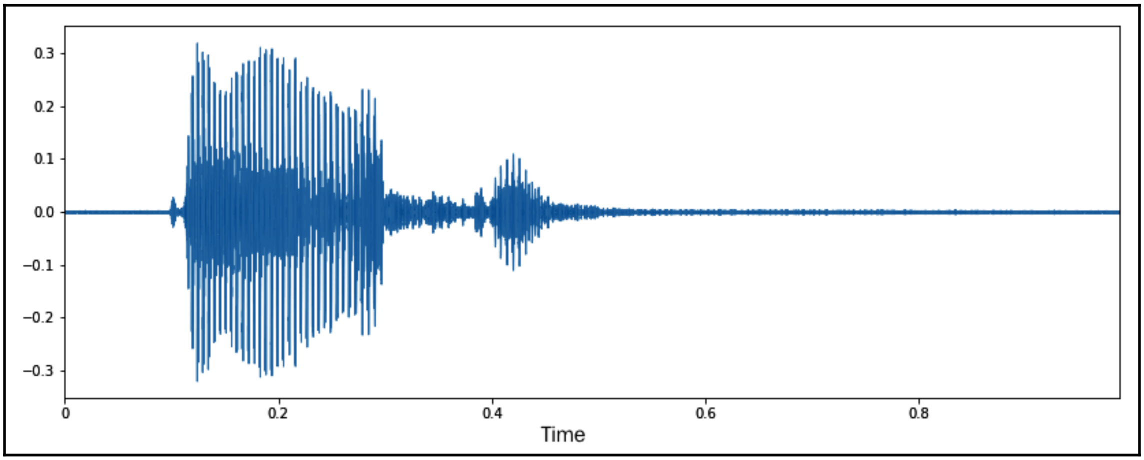

40
20
0
−20
−40
−40
−20
0
20
40

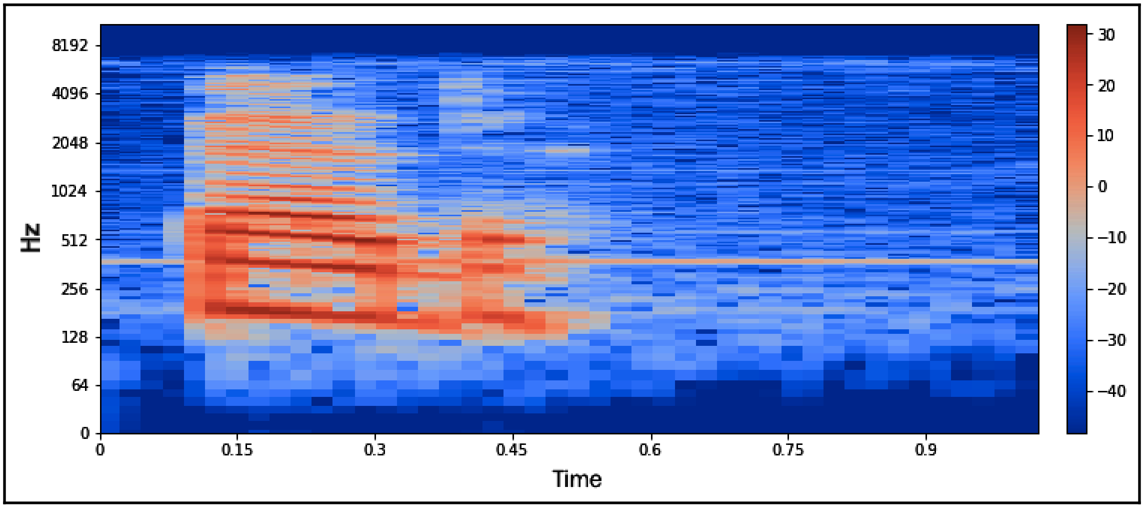

car 0.92728674
bench 0.9929489
0
50
100
150
200
250
300
350
400
0
100
200
300
400
500
600


person 0.9993374

Chapter 9: Deep Learning in Audio and Speech



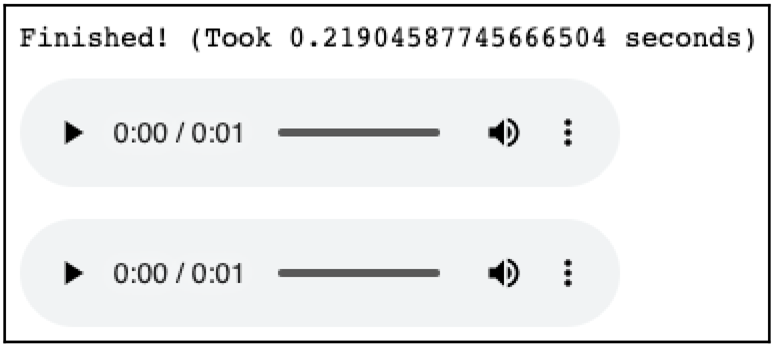




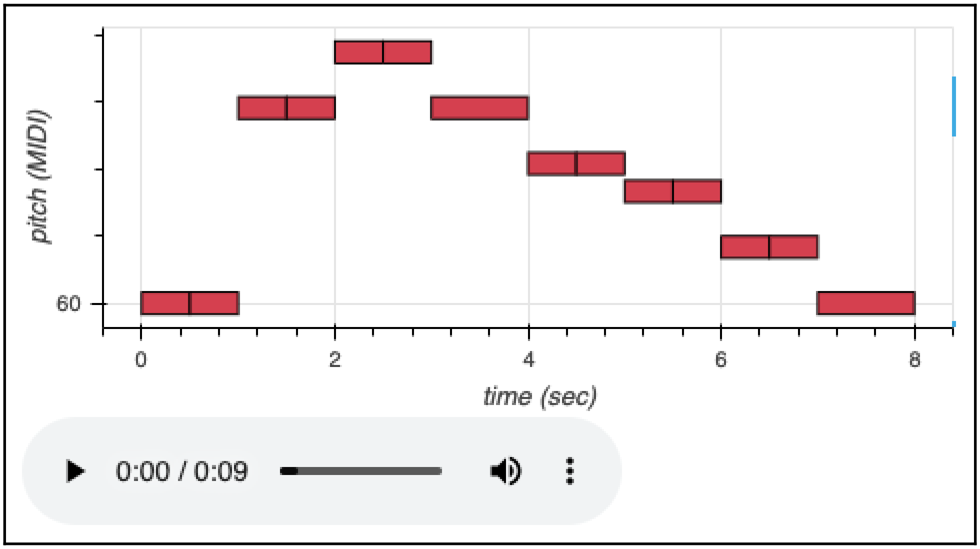
$$\mathcal{L}_{\text{att}}(A) = \mathbb{E}_{nt}[A_{nt}W_{nt}], \text{ where}$$

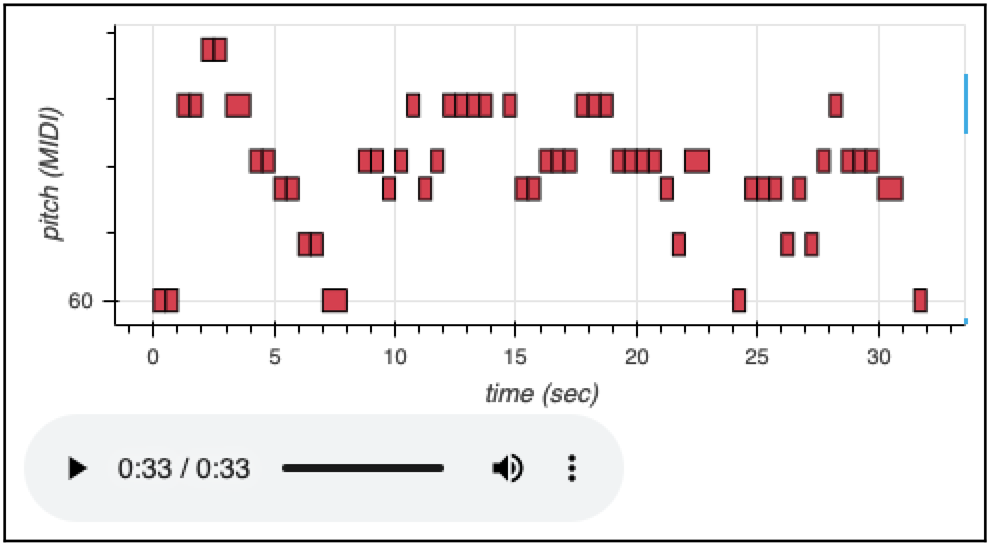
$$W_{nt} = 1 - \exp - \frac{(n/N - t/T)^2}{2g^2},$$

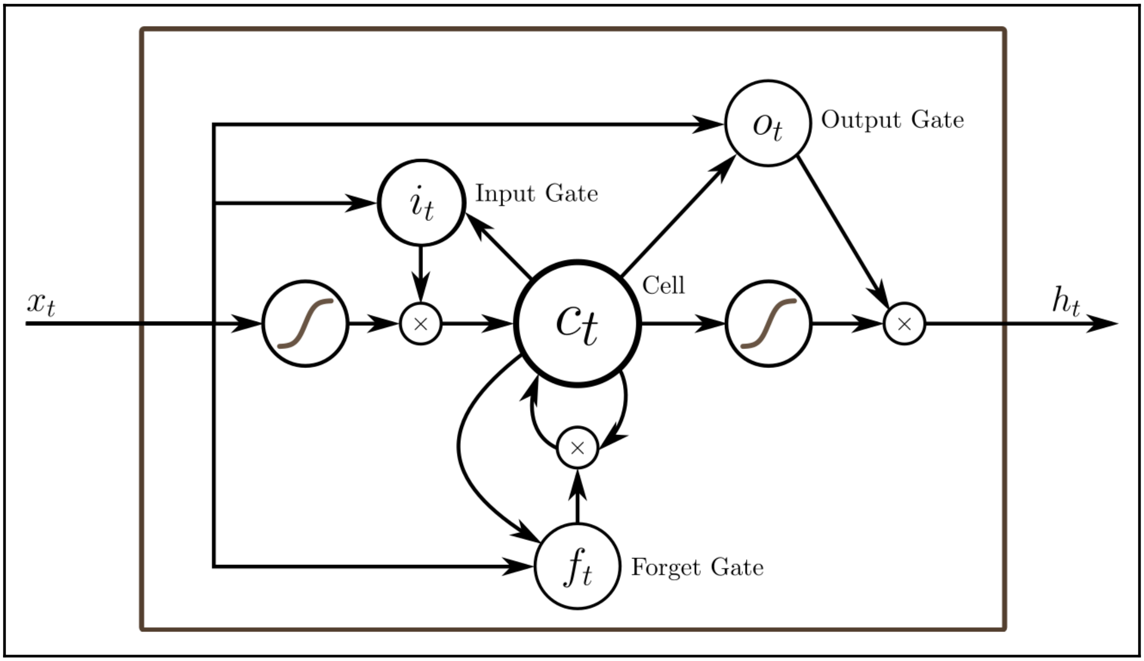
with parameter $g = 0.2$, N the number of characters, T the number of time frames, and $A \in \mathbb{R}^{NxT}$.



```
Finished! (Took 0.21904587745666504 seconds)
```




pitch (MIDI)
60
0
5
10
15
20
25
30
time (sec)
0:33 / 0:33

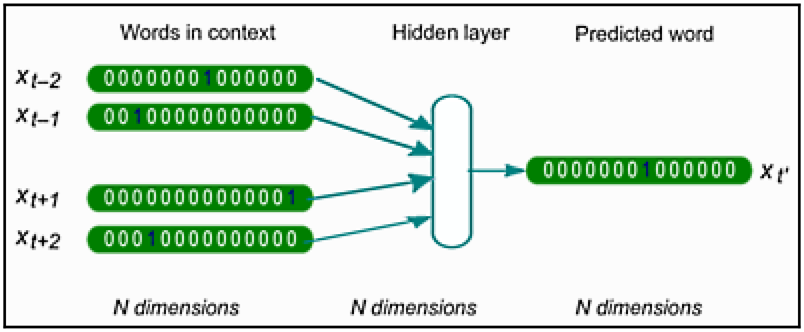

o_t
Output Gate
i_t
Input Gate
x_t
×
c_t
Cell
×
h_t
×
f_t
Forget Gate

$$u_i^t = v^T \tanh(W_1 h_i + W_2 c_t)$$

$$a_i^t = \mathrm{softmax}(u_i^t)$$

$$\hat{h}_t = \sum_{i=t-n}^{t-1} a_i^t h_i$$

over hidden states h and current RNN state c_t with model parameters v, W_1, and W_2.



$$\mathrm{king} - \mathrm{man} + \mathrm{woman} = \mathrm{queen}$$

$$\mathrm{man} - \mathrm{woman} \sim \mathrm{computer\ programmer} - \mathrm{homemaker}$$

$$1 \times N$$

$$N$$

$$|D| \times \boldsymbol{N}$$

$$D$$

$$|D|$$

$$D$$

$$\mathrm{tidf}(t, D) = (1 + \log f_{t,d}) \cdot \log \frac{N}{n_t},$$

$$f_{t,d}$$

$$t$$

d

n_t

t

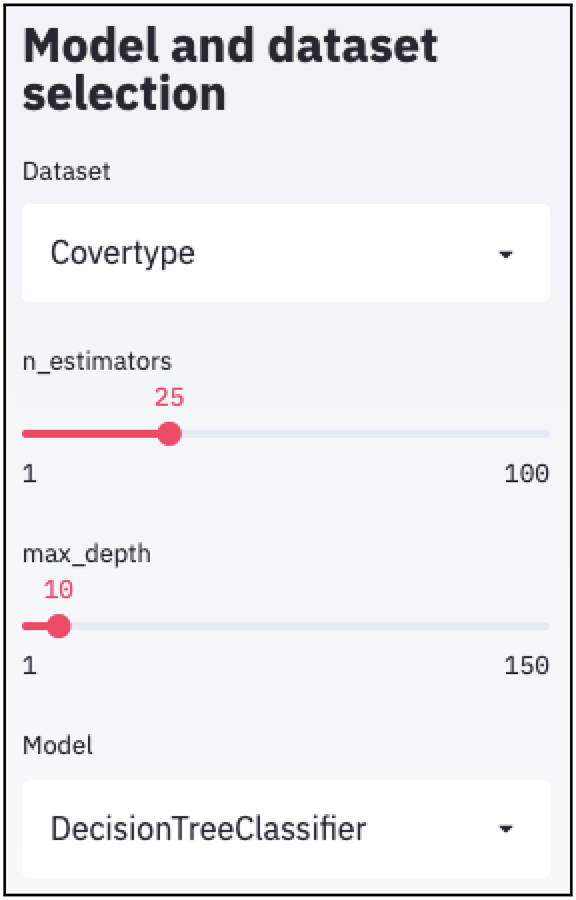
n_t

$$\sqrt{\text{hidden\_dim}}$$

$$\boldsymbol{n}$$

$$P(x_t | P_{t-1}, P_{t-2}, \ldots, P_{t-n})$$

Chapter 11: Artificial Intelligence in Production



Model and dataset selection

Dataset

Covertype

n_estimators

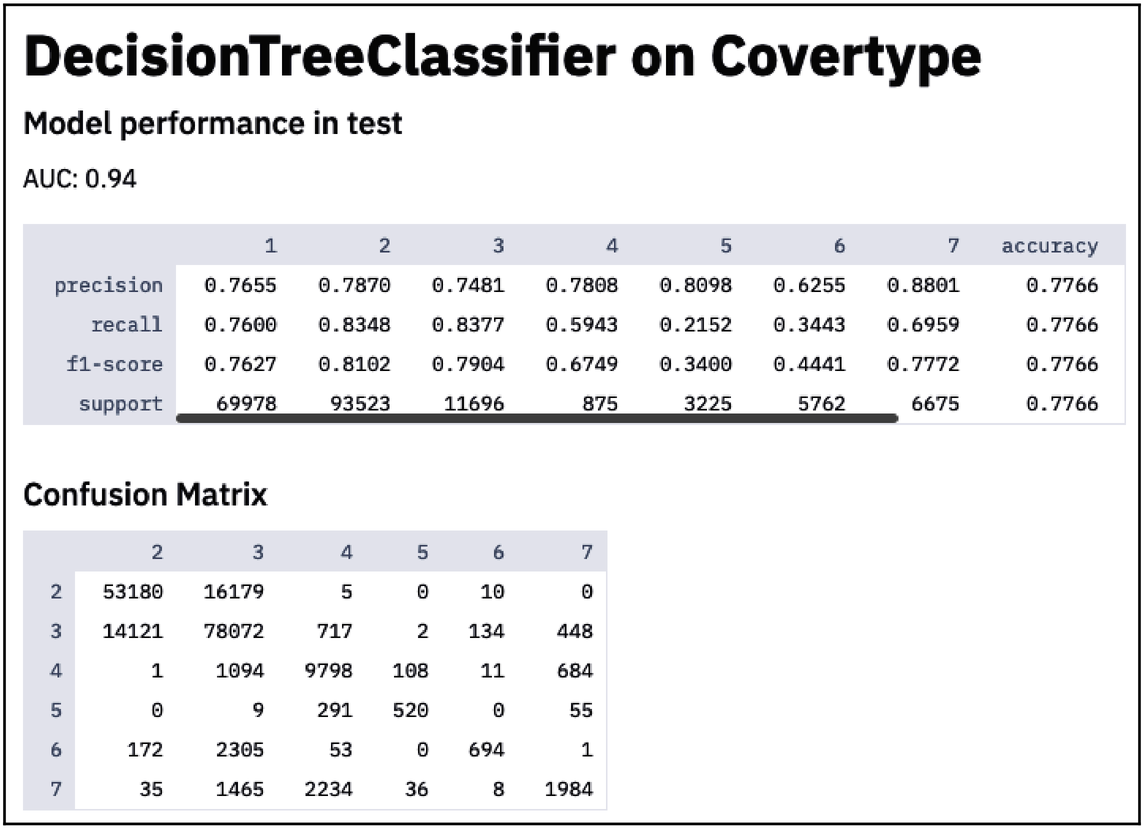
25

max_depth

10

Model

DecisionTreeClassifier



DecisionTreeClassifier on Covertype

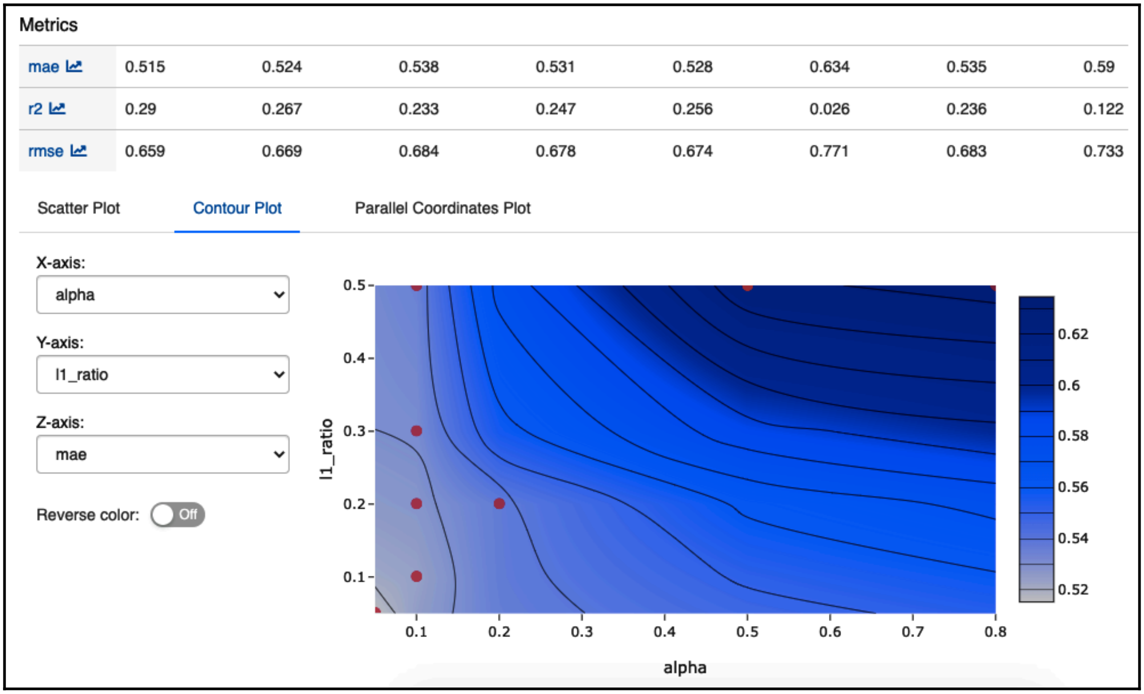
Model performance in test

AUC: 0.94

	1	2	3	4	5	6	7	accuracy
precision	0.7655	0.7870	0.7481	0.7808	0.8098	0.6255	0.8801	0.7766
recall	0.7600	0.8348	0.8377	0.5943	0.2152	0.3443	0.6959	0.7766
f1-score	0.7627	0.8102	0.7904	0.6749	0.3400	0.4441	0.7772	0.7766
support	69978	93523	11696	875	3225	5762	6675	0.7766

Confusion Matrix

	2	3	4	5	6	7
2	53180	16179	5	0	10	0
3	14121	78072	717	2	134	448
4	1	1094	9798	108	11	684
5	0	9	291	520	0	55
6	172	2305	53	0	694	1
7	35	1465	2234	36	8	1984



Metrics

mae	0.515	0.524	0.538	0.531	0.528	0.634	0.535	0.59
r2	0.29	0.267	0.233	0.247	0.256	0.026	0.236	0.122
rmse	0.659	0.669	0.684	0.678	0.674	0.771	0.683	0.733

X-axis: alpha

Y-axis: l1_ratio

Z-axis: mae

Reverse color: Off

$$Pr[M(D) \in S] \leq e^{\epsilon} Pr[M(D') \in S] + \delta, \text{ for } D', D \in \mathcal{D}.$$

$$M$$

$$D, D' \in \mathcal{D}$$

$$S$$

$$\mathcal{S}$$

$$\mathcal{N}(0,\sigma^2)$$

$$\mathcal{M}_\sigma(x) \mathrel{\hat{=}} \mathrm{argmax}_{\mathrm{i}}\{\mathrm{n}_{\mathrm{i}}(\mathrm{x}) + \mathcal{N}(0,\sigma^2)\},$$