Data Mining In Action

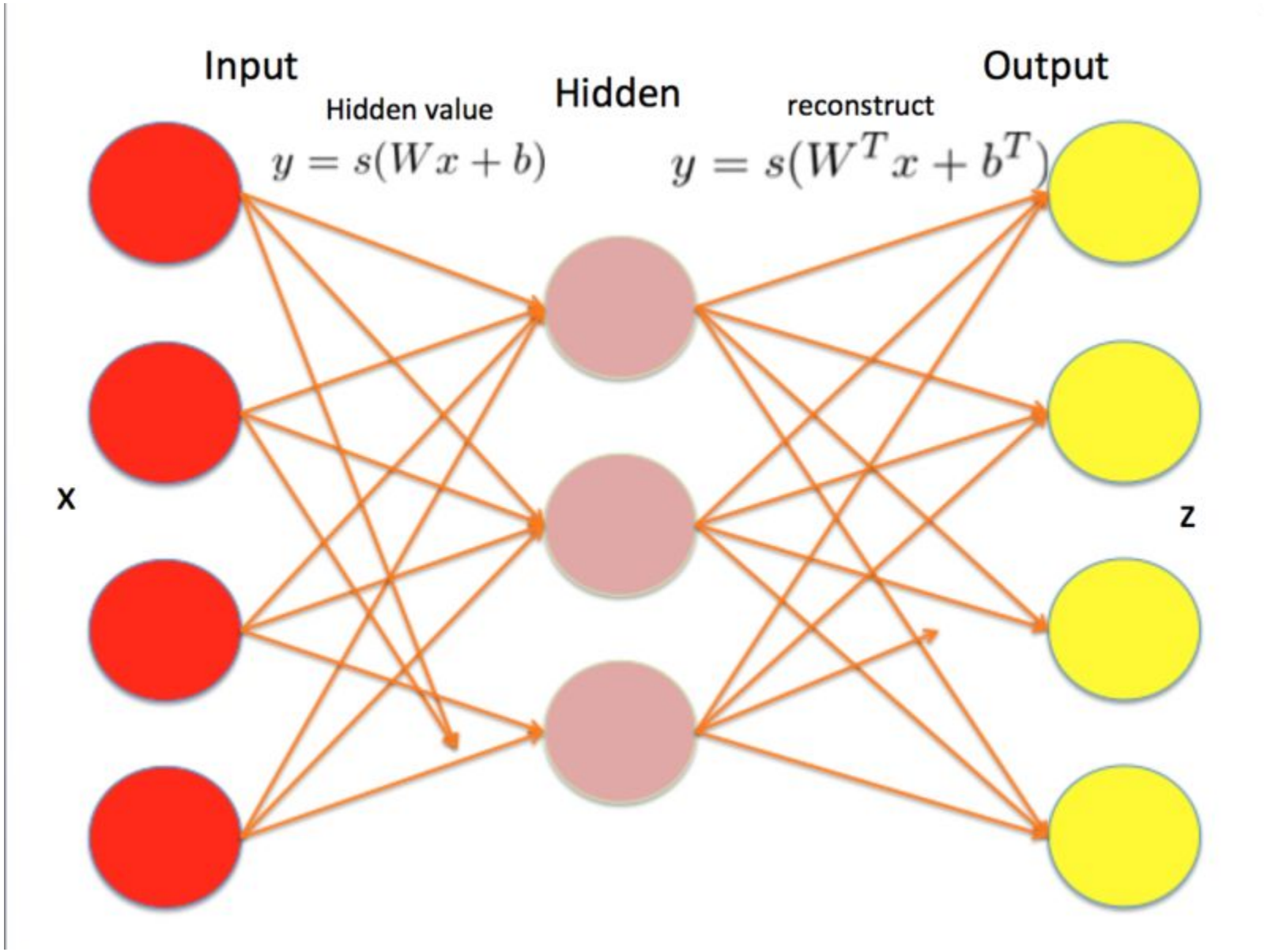
Тренды, разбор дз

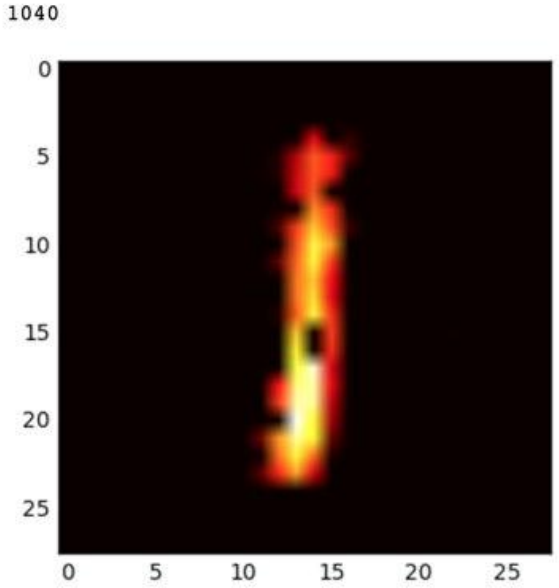
"Your recent Amazon purchases, Tweet score and location history makes you 23.5% welcome here."

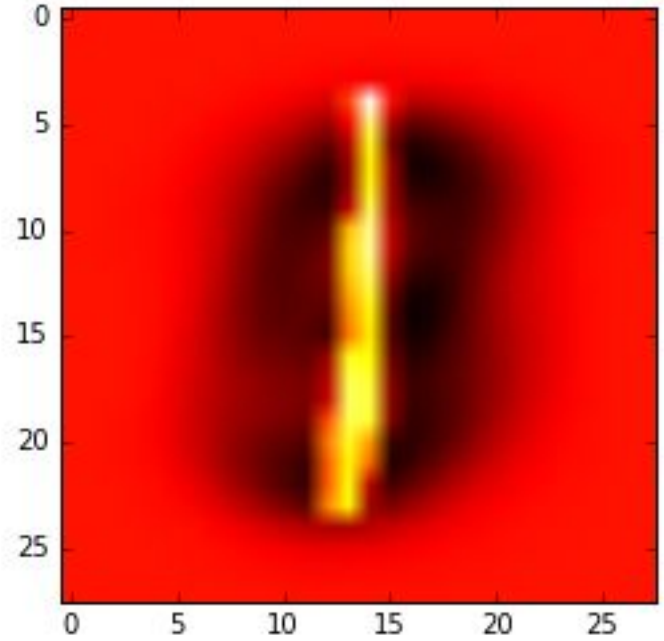
Ашуха Арсений
Дмитрий Персиянов
Антон Панкратов

https://vk.com/data_mining_in_action

15.11.2016

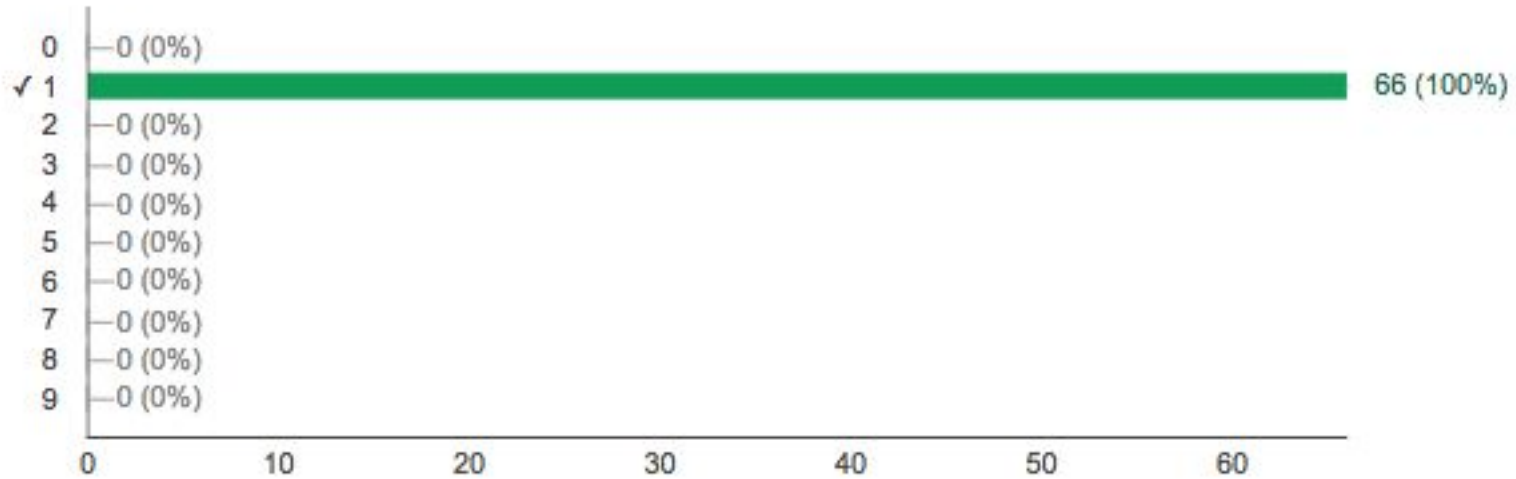

Input
Hidden value
$y = s(Wx + b)$
Hidden
reconstruct
$y = s(W^T x + b^T)$
Output
x
z

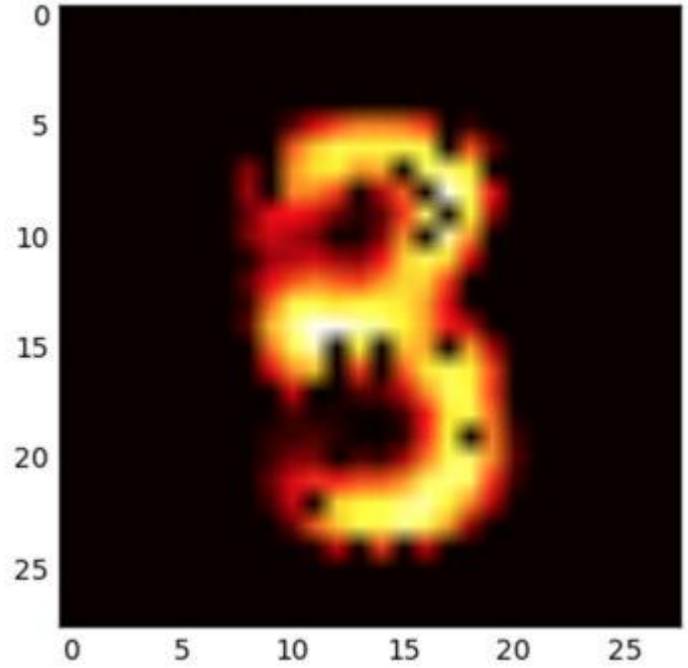


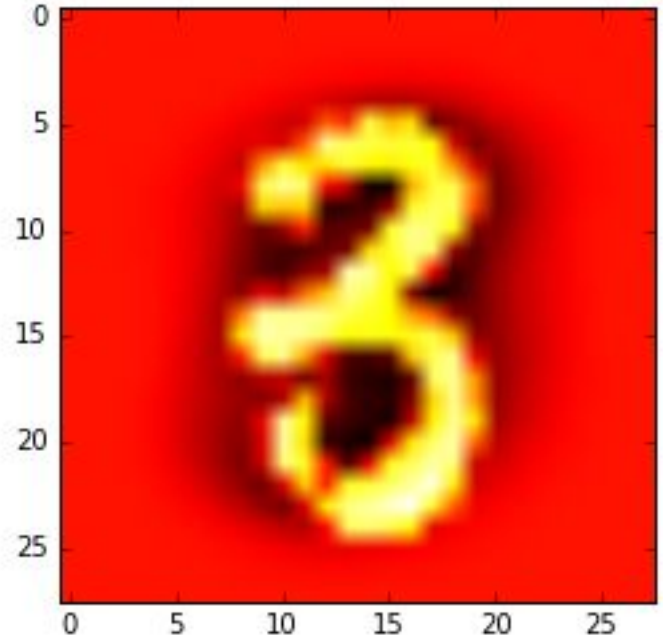


Лейбл четвертой цифры (y[1040])

66 / 66 correct responses

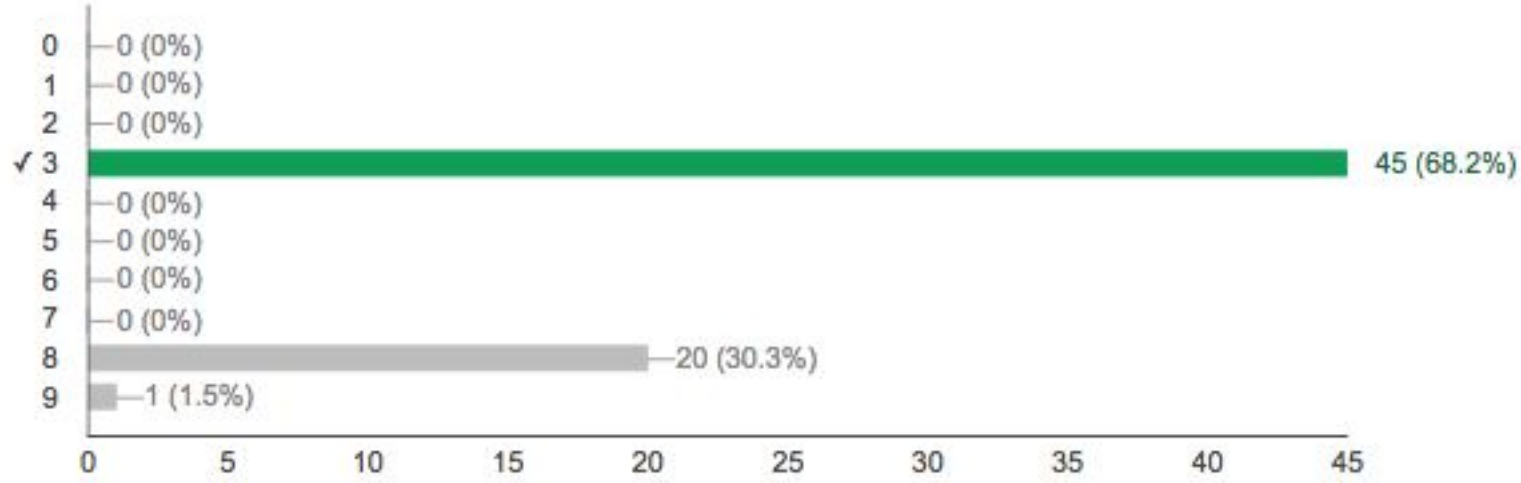




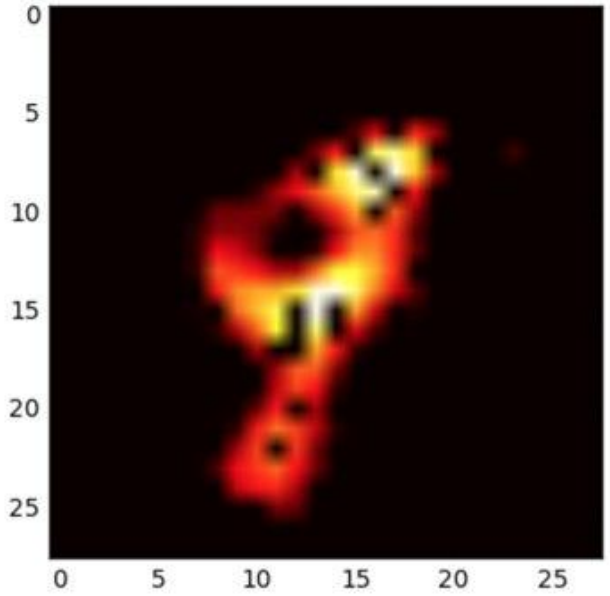


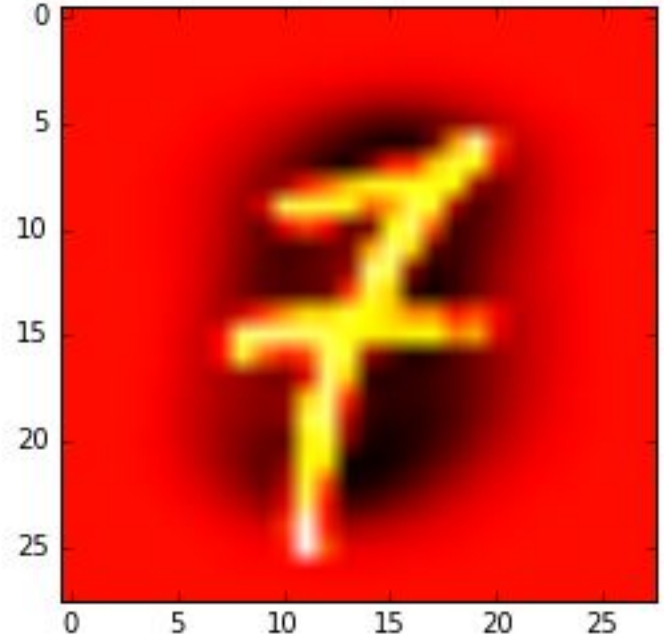
Лейбл второй цифры (y[2853])

45 / 66 correct responses



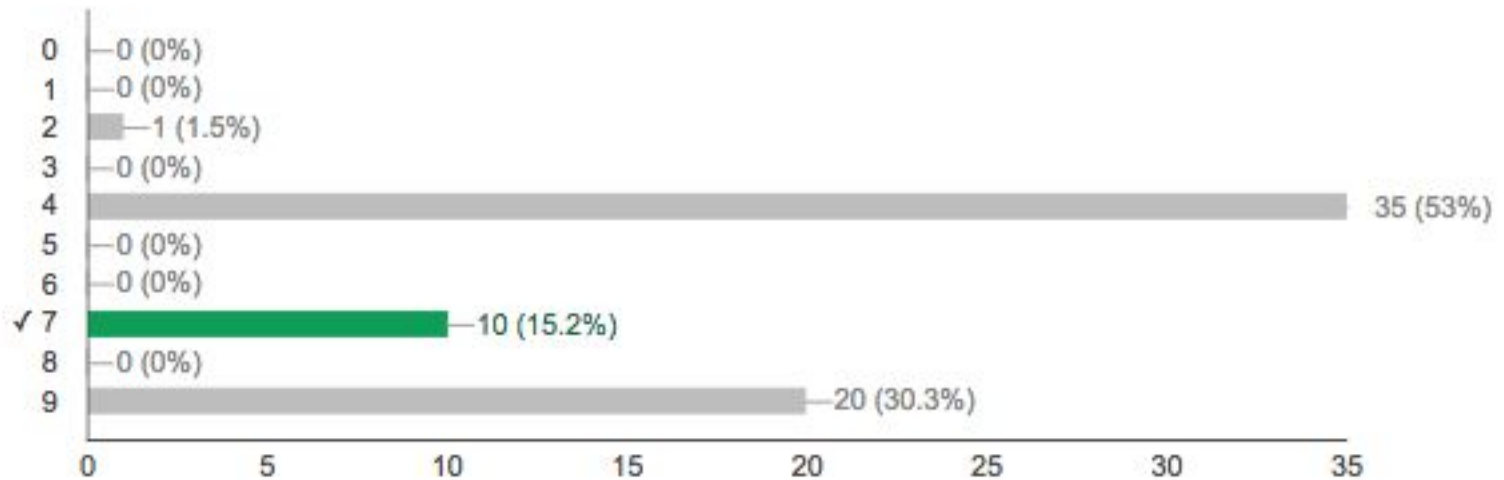
3136



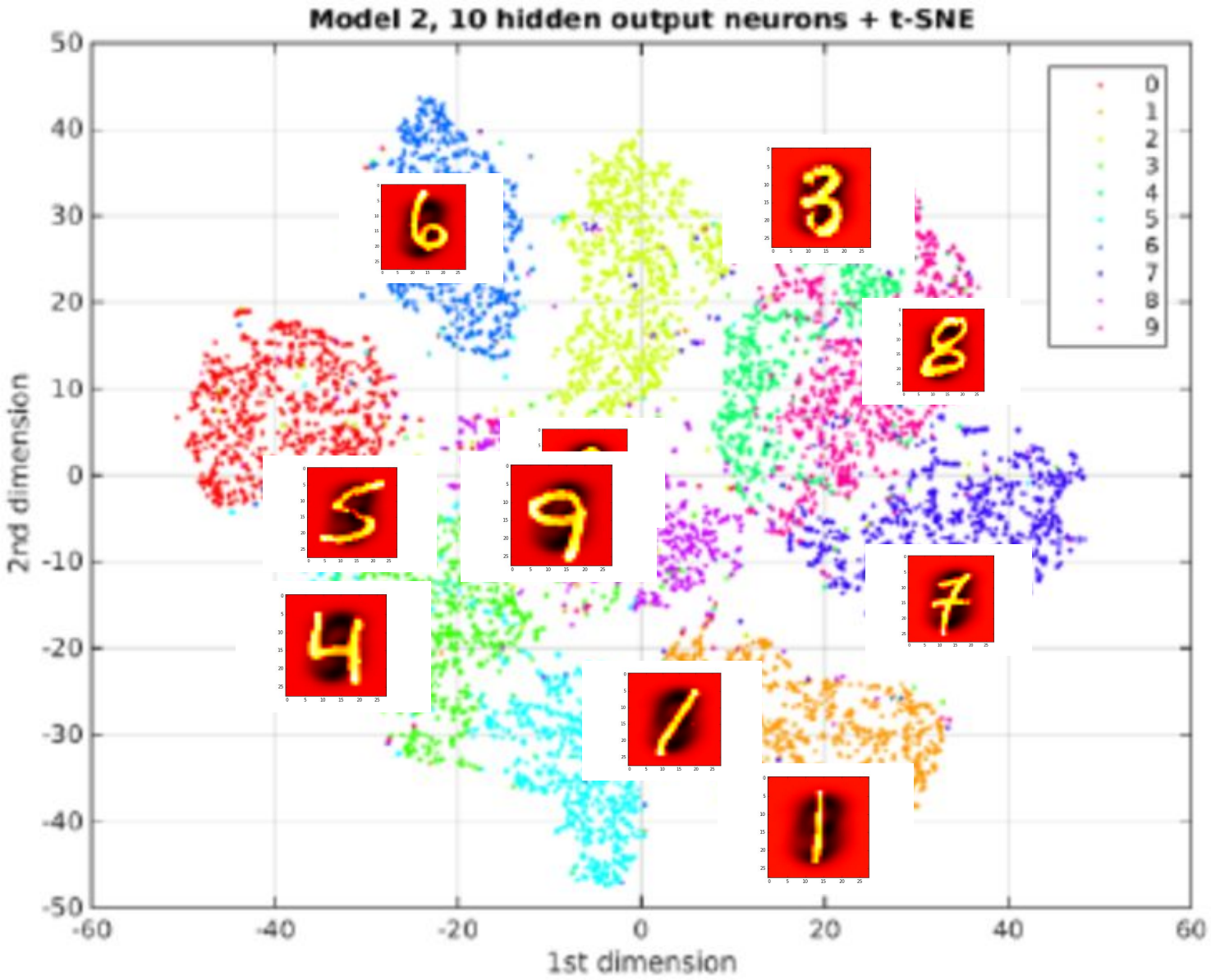


Лейбл третей цифры (y[3136])

10 / 66 correct responses



TSNE


Model 2, 10 hidden output neurons + t-SNE
2nd dimension
1st dimension
50
40
30
20
10
0
-10
-20
-30
-40
-50
-60
-40
-20
0
20
40
60
0
1
2
3
4
5
6
7
8
9

Cheating

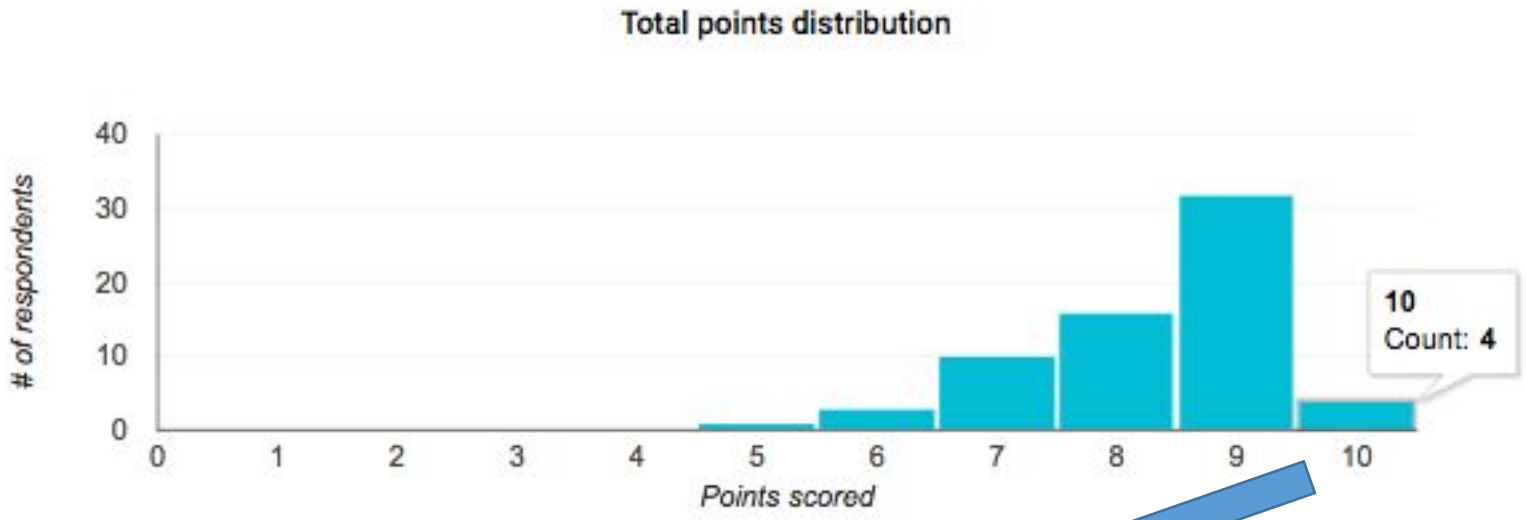
In [8]:
```
from sklearn.svm import SVC
```

In [9]:
```
X_train_mid, _, X_test_mid = np.load('X_train_mid.npy'), X_train, np.load('X_test_mid.npy')
clf = SVC().fit(X_train_mid, y_train)
```

In [10]:
```
print(clf.predict(X_test_mid[idx]))
print(y_test[idx])
print ((clf.predict(X_test_mid[idx]) == y_test[idx]).sum())
```

```
[9 3 4 1 6 5 4 2 1 9]
[9 3 7 1 6 5 4 8 1 9]
8
```

Average	Median	Range
8.32 / 10 points	9 / 10 points	5 - 10 points



- Анна Кобякова
- Игорь Козловский
- Александр Селютин
- Дмитрий Нехаев