CSE250b_HW3

Kai Zhou

February 2, 2016

1 Bivariate Gaussians

1.1 a

According to the problemm, we have

	μ	σ
x	2	1
y	2	0.5

because $corr(X,Y) = \frac{cov(X,Y)}{std(X)std(Y)}$, $corr(X,Y) = -0.5$, $std(X) = 1$ and $std(Y) = 0.5$ we can get the parameter is:

$$\Sigma = \begin{pmatrix} 1 & \text{-025} \\ \text{-0.25} & 0.25 \end{pmatrix}$$

1.2 b

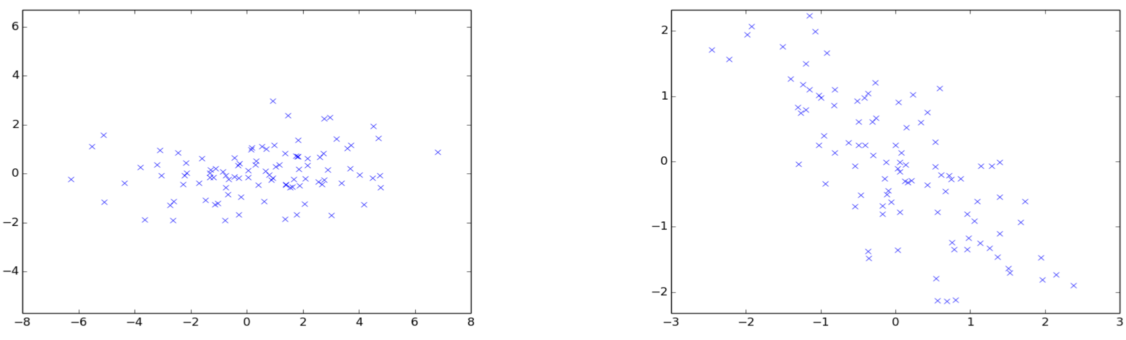
According to the problemm, we have

	μ	σ
x	1	1
y	1	1

so $cov(X,Y) = 0$. The parameter is:

$$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

2 More bivariate Gaussians

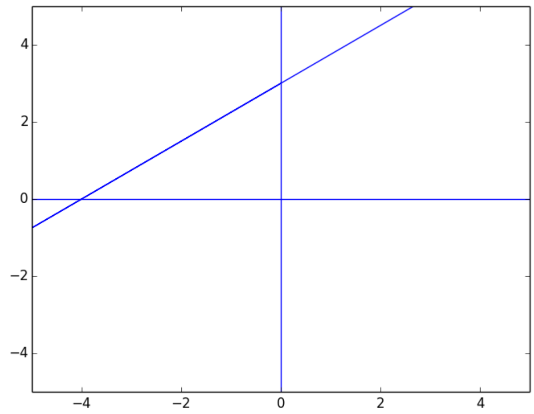


(a) 2a

(b) 2b

Figure 1: Bivariate Gaussian Plot

3 Linear classification



The left up side is the positive side.

4 Eigendecomposition

4.1 a

If there is no ZERO eigenvalue, then the matrix is invertible.
Proof: if $\lambda = 0$, then we have $\Sigma \cdot v = \lambda \cdot v = 0$. If Σ is invertible, then $\Sigma \cdot v = 0$ iff $v = 0$. But according to the definition of eigenvalue, $v \neq 0$. Here is a contradiction. So if there is no zero eigenvalue, the matrix is invertible.

4.2 b

$$\Sigma \cdot v = \lambda \cdot v$$

$$\begin{aligned}(\Sigma + c \cdot I) \cdot v &= \Sigma \cdot v + c \cdot I \cdot v \\ &= \Sigma \cdot v + c \cdot I \cdot v \\ &= \lambda \cdot v + c \cdot v \\ &= (\lambda + c) \cdot v\end{aligned}$$

So the eigenvalues are $\lambda + c$ and the eigenvectors stay are the sane as the eigenvectors of Σ.

4.3 c

$$\begin{aligned}\Sigma \cdot v &= \lambda \cdot v \\ \Sigma^{-1} \cdot \Sigma \cdot v &= \lambda \cdot \Sigma^{-1} \cdot v \\ v &= \lambda \cdot \Sigma^{-1} \cdot v \\ \frac{1}{\lambda} \cdot v &= \Sigma^{-1} \cdot v\end{aligned}$$

So the eigenvalues are $\frac{1}{\lambda}$ and the eigenvectors stay are the sane as the eigenvectors of Σ.

5 Handwritten digit recognition using a Gaussian generative model

5.1 Pseudocode

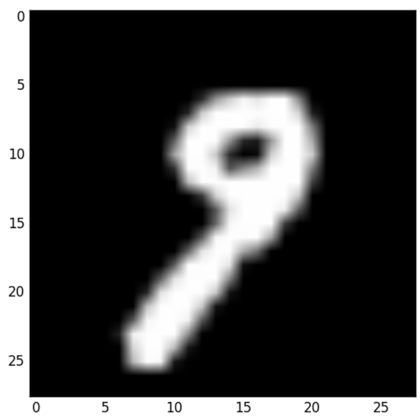
Data: train_data, train_label, test_data, test_label
Result: parameter c which is used to smooth the covariance
Randomly choose 10000 number from 0 - 59999 as the validation data indexs;
We create the validation data and validation labels according to the indexs and the remaining as the new train data;
Divide the train data to 10 classes according to their labels;
Create the mean array and covariance matrixs for the 10 classes;
for *c from 0.01 to 10000* **do**
    Smooth the covariance by adding in cI;
    Train 10 classifiers of the Gaussian generative model;
    Apply the classifiers to the validation data to get a error rate;
    if *error rate is the local minimum* **then**
        return c
    end
end

Algorithm 1: Pseudocode of training procedure

5.2 error rate on test set

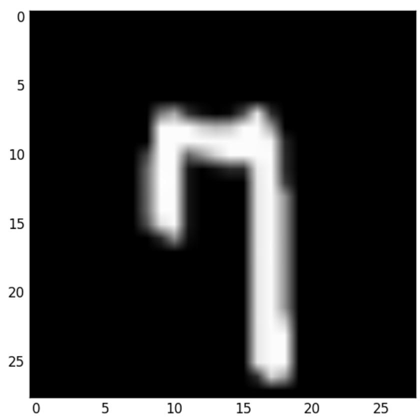
Apply the algorithm above, we find the local minimun of error rate on validation data is when $c = 3000$. Using $c = 3000$, the error rate on test data is 4.32%

5.3 posterior probabilities



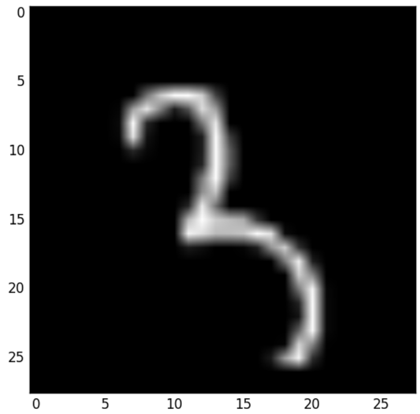
Label : 9
Prediction Label : 7

0	1	2	3	4
3.84e-58	1.88e-43	1.57e-34	8.41e-32	7.14e-30
5	6	7	8	9
4.91e-48	5.46e-89	0.99985	3.20e-14	0.00015



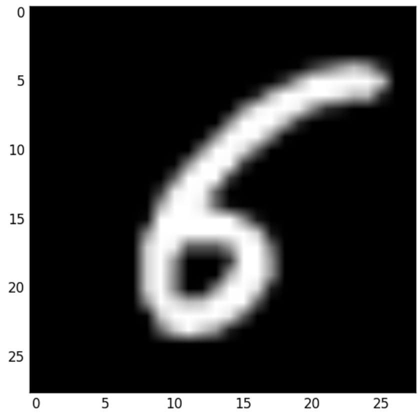
Label :7
Prediction Label : 9

0	1	2	3	4
5.41e-69	1.02e-89	3.53e-61	4.21e-42	8.16e-19
5	6	7	8	9
1.47e-38	7.94e-94	0.30263	6.97e-37	0.69736



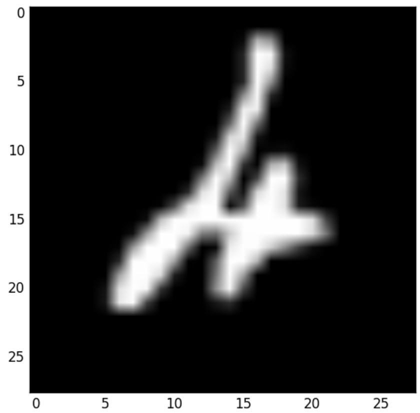
Label :3
Prediction Label : 4

0	1	2	3	4
3.49e-31	8.60e-20	6.15e-07	0.01083	0.71794
5	6	7	8	9
0.01112	3.85e-22	6.33e-15	0.25865	0.00144



Label :6
Prediction Label : 5

0	1	2	3	4
2.85e-35	4.97e-66	2.76e-46	2.98e-35	7.98e-43
5	6	7	8	9
0.99999	4.10e-12	2.20e-96	5.35e-17	1.21e-75



Label :4
Prediction Label : 6

0	1	2	3	4
7.94e-31	7.05e-41	0.00061	5.25e-21	2.04e-20
5	6	7	8	9
9.66e-25	0.99938	6.15e-53	4.95e-22	3.82e-62

6 A classifier for MNIST that occasionally abstains

6.1 description

6.2 Pseducode

6.3 two curves