CSE 250B: Homework 2

1. Error rate of 1-NN classifier

(a) the example is the following:
(1,1), (3,1), (2, 1+$\sqrt{3}$) and they belong to cluster 0, 0, 1 respectively.
(b) When R^* is zero, R $\rightarrow 2R^*(1 - R^*) = 0$. Therefore, $1 - NN$ classification is consistent when Bayes risk R^* is zero.

2. Bayes optimality in a multi-class setting

$$h^*(x) = arg_i max(\eta_i(x)) \tag{1}$$

3. Classification with an abstain option

For one specific x, we consider the cases:
(1) if the label is 0, then the cost of predicting 1 is $\eta(x)$ and the cost of predicting abstain is θ.
(2) if the label is 1, then the cost of predicting 0 is $1 - \eta(x)$ and the cost of predicting abstain is θ.
Hence, the classifier that has minimum expected cost is the following:

$$h(x) = \begin{cases} 1 & 1 - \eta(x) < \theta \\ 0 & \eta(x) < \theta \\ abstain & otherwise \end{cases} \tag{2}$$

4. The statistical learning assumption

(a) This case violates the statistical assumption because η changes.

1. People's taste is not static. Different times have different trends and genres in music. What became a hit doesn't necessarily mean that it would succeed again in the future.
2. Success cannot be duplicated. People don't like similar stuff. When some kind of music is repeated too many times, people would get tired of it.

(b) The statistical assumption would hold in this case.

(c) This case violates the statistical assumption because μ changes.

1. This online dating site worked well on the west coast doesn't necessarily mean that it would work well nationwide. This is because people living in west coast cannot represent all people in the country. People may think different, act different because they live and grow up in different environments.

5. Conditional probability

(a)

$$\begin{aligned} Pr(happy \mid talksalittle) &= \frac{Pr(talksalittle \mid happy)Pr(happy)}{Pr(talksalittle)} \\ &= \frac{\frac{1}{6} \times \frac{3}{4}}{\frac{1}{6} \times \frac{3}{4} + \frac{1}{6} \times \frac{1}{4}} = \frac{3}{4} \end{aligned} \tag{3}$$

$$Pr(sad \mid talksalittle) = \frac{1}{4} \tag{4}$$

Hence, he is most likely to be happy.

(b) The probability of being incorrect is $\frac{1}{4}$.

6. Bayes optimal classifier

$$\begin{aligned} \eta_1(x) = Pr(Y = 1 \mid X = x) &= \frac{Pr(X = x \mid Y = 1)Pr(Y = 1)}{Pr(X = x)} \\ &= \begin{cases} \frac{7}{13} & -1 < x < 0 \\ \frac{1}{11} & 0 \leq x < 1 \end{cases} \end{aligned} \tag{5}$$

$$\begin{aligned} \eta_2(x) = Pr(Y = 2 \mid X = x) &= \frac{Pr(X = x \mid Y = 2)Pr(Y = 2)}{Pr(X = x)} \\ &= \begin{cases} 0 & -1 < x < 0 \\ \frac{4}{11} & 0 \leq x < 1 \end{cases} \end{aligned} \tag{6}$$

$$\begin{aligned} \eta_3(x) = Pr(Y = 3 \mid X = x) &= \frac{Pr(X = x \mid Y = 3)Pr(Y = 3)}{Pr(X = x)} \\ &= \begin{cases} \frac{6}{13} & -1 < x < 0 \\ \frac{6}{11} & 0 \leq x < 1 \end{cases} \end{aligned} \tag{7}$$

$$h^*(x) = \begin{cases} 1 & -1 < x < 0 \\ 3 & 0 \leq x < 1 \end{cases} \tag{8}$$

$$\begin{aligned} R^* = Pr(h^*(x) \neq y) = Pr(h^*(x) = 1, y \neq 1) + Pr(h^*(x) = 3, y \neq 3) \\ = Pr(y \neq 1 \mid h^*(x) = 1)Pr(h^*(x) = 1) \\ + Pr(y \neq 3 \mid h^*(x) = 3)Pr(h^*(x) = 3) \\ = \frac{6}{13} \times \frac{1}{2} + \frac{5}{11} \times \frac{1}{2} = 0.458 \end{aligned} \tag{9}$$

7. Covariance and correlation

(a)

$$\mu_x = 0, \sigma_x = 10 \tag{10}$$

Since $Y = 2X$, we have

$$\mu_y = 0, \sigma_y = 20 \tag{11}$$

Since $Var(X) = \sigma_x^2 = E(X^2) - E^2(X) = 100$, we have $E(X^2) = 100$. Thus $Cov(X,Y) = E(XY) - E(X)E(Y) = 2E(X^2) = 200$.

Hence, the covariance matrix is $\begin{pmatrix} 100 & 200 \\ 200 & 400 \end{pmatrix}$.

(b) correlation coefficient $\rho = \frac{200}{\sigma_x \sigma_y} = 1$.

8. Bivariate Gaussians

(a)

$$\mu_x = 2, \sigma_x = 1 \tag{12}$$

$$\mu_y = 2, \sigma_y = 0.5 \tag{13}$$

$Cov(X,Y) = corr \times \sigma_x \times \sigma_y = -0.25$. Hence, $\mu = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$, covariance matrix $\sum = \begin{pmatrix} 1 & -0.25 \\ -0.25 & 0.25 \end{pmatrix}$.

And we can know that $Inv(\sum) = \frac{4}{3} \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix}$.

$$\begin{aligned} p(x,y) &= \frac{1}{2\pi \times \frac{\sqrt{3}}{4}} exp(-\frac{1}{2}[x-2, y-2]Inv(\sum)[x-2, y-2]^T) \\ &= \frac{2}{\sqrt{3}\pi} exp(\frac{4}{3}(x^2 + 4y^2 - 8x - 18y + 2xy + 28)) \end{aligned} \tag{14}$$

(b) $\mu = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, covariance matrix $\sum = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$.

correlation $= \frac{Cov(X,Y)}{\sigma_x \sigma_y} = 1$.

In this case, covariance matrix doesn't have inverse matrix.

$$p(x,y) = \begin{cases} \frac{1}{\sqrt{2\pi}} exp(-\frac{(x-1)^2}{2}) & x = y \\ 0 & x \neq y \end{cases} \tag{15}$$

9. More bivariate Gaussians

(a) The contour is the following:

The figure containing 100 random samples is the following:

(b) The contour is the following:

The figure containing 100 random samples is the following:

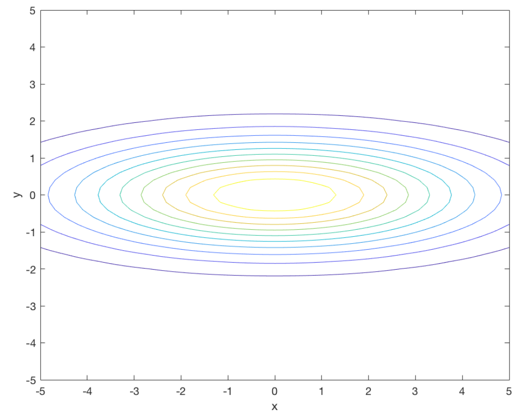

Figure 1: contour

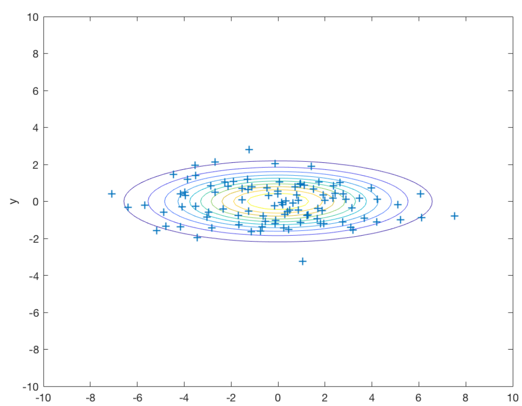

Figure 2: 100 random samples

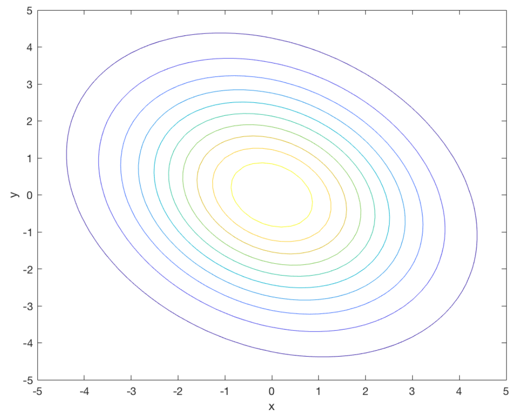

Figure 3: contour

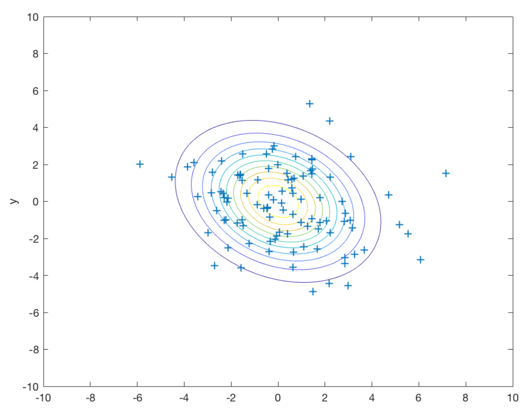

Figure 4: 100 random samples

10. Qualitative appraisal of Gaussian parameters

(a) If the two variables are negatively correlated, $b < 0, a \geq 0, c \geq 0$.
Also, to ensure that $\mid \sum \mid > 0$, we have

$$ac - b^2 > 0 \Rightarrow 0 > b > -\sqrt{ac} \tag{16}$$

(b) If the two variables are uncorrelated, $b = 0, a \geq 0, c \geq 0$.
(c) If one variable is a linear function of the other: without loss of generality, let $Y = kX + b$ where k, b are constant values. Thus we have

$$a = Var(X), c = Var(Y) = k^2 a \tag{17}$$

$$b = Cov(X, Y) = Cov(X, kX + b) = kVar(X) = ka \tag{18}$$

$$c \geq 0 \tag{19}$$

(d) If one of the variables is a constant:
if X is a constant (denoted as k), we have

$$a = Var(X) = 0, b = Cov(X, Y) = E(kY) - E(k)E(Y) = 0, c = Var(Y) \tag{20}$$

similarly, if Y is a constant (denoted as k), we have

$$a = Var(X), b = Cov(X, Y) = E(kX) - E(k)E(X) = 0, c = Var(Y) = 0 \tag{21}$$