UNIVERSITY OF TEXAS AT AUSTIN

Homework Assignment 8

Correlation. Bivariate normal. LDA. QDA.

Please, provide your **complete solutions** to the following problems. Final answers only, even if correct will earn zero points for those problems.

Problem 8.1. (10 points) Let X and Y be two random variables with finite first and second moments. You know that the correlation between X and Y equals 1, i.e., the two random variables are perfectly positively correlated. Prove that each can be expressed as a linear transform of the other. More precisely, prove that there exist α and β such that $Y = \alpha X + \beta$.

Solution: First, define a random variable $W = Y - \frac{\sigma_Y}{\sigma_X}X$, where σ_X and σ_Y stand for the standard deviations of X and Y, respectively. Let's find the variance of W. By the definition of W, we have that

$$\mathrm{Var}[W] = \mathrm{Var}[Y - \frac{\sigma_Y}{\sigma_X}X].$$

Using the covariance formula, we get

$$\begin{aligned}\mathrm{Var}[W] &= \mathrm{Var}[Y] - 2\frac{\sigma_Y}{\sigma_X}\,\mathrm{Cov}[X,Y] + \frac{\sigma_Y^2}{\sigma_X^2}\,\mathrm{Var}[X]\\ &= \sigma_Y^2 - 2\frac{\sigma_Y}{\sigma_X}\sigma_X\sigma_Y\rho(X,Y) + \sigma_Y^2\end{aligned}$$

where $\rho(X,Y)$ stands for the correlation coefficient between X and Y. Simplifying the above expression and using the given fact that $\rho(X,Y) = 1$, we obtain $\mathrm{Var}[W] = 0$. Hence, W is constant with probability 1. Let's call this constant α. We now know that

$$\alpha = Y - \frac{\sigma_Y}{\sigma_X}X \quad \Leftrightarrow \quad Y = \alpha + \beta X$$

with β defined as $\frac{\sigma_Y}{\sigma_X}$.

Problem 8.2. (10 points) *Source: "Probability" by Jim Pitman.*
Data from a large population indicate that the heights of mothers and daughters in this population follow the bivariate normal distribution with correlation 0.5. Both variables have mean 5 feet 4 inches, and standard deviation 2 inches. Among the daughters of above average height, what percent were shorter than their mothers?

Solution: Let the random variable X be the daughters' heights in standard units and let Y be the mothers' heights in standard units. We are looking for the following probability:

$$\mathbb{P}[X < Y \,|\, X > 0]$$

As we learned in class, Y can be expressed as $Y = \frac{1}{2}X + \frac{\sqrt{3}}{2}Z$ for $Z \sim N(0,1)$ independent from X. Then, our conditional probability becomes

$$\mathbb{P}\left[X < \tfrac{1}{2}X + \tfrac{\sqrt{3}}{2}Z \,|\, X > 0\right] = \mathbb{P}[X < \sqrt{3}Z \,|\, X > 0].$$

By the definition of conditional probability, the above equals

$$\frac{\mathbb{P}[0 < X < \sqrt{3}Z]}{\mathbb{P}[X > 0]}.$$

Obviously, $\mathbb{P}[X > 0] = \frac{1}{2}$. On the other hand, the event in the probability in the numerator is a cone (wedge) in the plane occupying 1/6 of the entire plane. Due to the radial symmetry of the joint distribution of (X, Z), the final answer is 1/3.

Problem 8.3. (20 points) Solve problem **4.8.3** (page 189) from the textbook.

Solution: This derivation is completely analogous to the one that was done in class in the homogeneous case, i.e., for LDA. Let us start with the expression for the posterior probability of being in class $k = 1, 2, \dots, K$.

$$p_k(x) = \frac{\pi_k \frac{1}{\sigma_k\sqrt{2\pi}} \exp(-\frac{1}{2\sigma_k^2}(x-\mu_k)^2)}{\sum_{l=1}^k \pi_\ell \frac{1}{\sigma_\ell\sqrt{2\pi}} \exp(-\frac{1}{2\sigma_\ell^2}(x-\mu_\ell)^2)}.$$

That's the probability that we are seeking to maximize. Note that the denominator does not depend on k. Using the fact that ln is an increasing function, we obtain that the above maximization problem is equivalent to

$$\ln(p_k(x)) = \ln(\pi_k) + \ln\left(\frac{1}{\sigma_k\sqrt{2\pi}}\right) - \frac{1}{2\sigma_k^2}(x-\mu_k)^2 \to \max.$$

The above is, in turn, equivalent to

$$\ln(p_k(x)) = \ln(\pi_k) - \frac{1}{2}\ln(2\pi) - \ln(\sigma_k) - \frac{x^2}{2\sigma_k^2} + \frac{2x\mu_k}{2\sigma_k^2} - \frac{\mu_k^2}{\sigma_k^2} \to \max.$$

Since the quadratic term x^2 is divided by σ_k^2 which **depends on** k, unlike in the case of $\sigma_1 = \dots = \sigma_K$, it remains in the classification criterion.

Problem 8.4. $(3+3+2+2 = 10$ points) Solve problem **4.8.5** (page 190) from the textbook.

Solution:

(a) QDA will always work better on the training set (as the more flexible model). However, with a truly linear boundary it will overfit and we can suspect that it will be worse on the testing set.

(b) QDA will always work better on the training set (as the more flexible model). With a truly non-linear boundary we can expect that it will work better on the testing set than the LDA.

(c) With more data points, the more granular QDA should yield improved accuracy. Here is a quote from page 153 in the textbook: *"Roughly speaking, LDA tends to be a better bet than QDA if there are relatively few training observations and so reducing variance is crucial. In contrast, QDA is recommended if the training set is very large, so that the variance of the classifier is not a major concern, or if the assumption of a common covariance matrix for the K classes is clearly untenable."*

(d) **FALSE**. As mentioned in the response to part (a), there is a danger of overfitting.