CS 189 Introduction to Machine Learning
Fall 2017 DIS4

1 Multivariate Gaussians: A review

(a) Consider a two dimensional random variable $Z \in \mathbb{R}^2$. In order for the random variable to be jointly Gaussian, a necessary and sufficient condition is that

- Z_1 and Z_2 are each marginally Gaussian, and
- $Z_1 | Z_2 = z$ is Gaussian, and $Z_2 | Z_1 = z$ is Gaussian.

A second characterization of a jointly Gaussian RV Z is that it can be written as $Z = AX$, where X is a collection of i.i.d. standard normal RVs and $A \in \mathbb{R}^{2\times 2}$ is a matrix.

Let X_1 and X_2 be i.i.d. standard normal RVs. Let U denote a random variable uniformly distributed on $\{-1, 1\}$, independent of everything else. Verify if the conditions of the first characterization hold for the following random variables, and calculate the covariance matrix Σ_Z.

- $Z_1 = X_1$ and $Z_2 = X_2$.
- $Z_1 = X_1$ and $Z_2 = X_1 + X_2$. (Use the second characterization to argue joint Gaussianity.)
- $Z_1 = X_1$ and $Z_2 = -X_1$.
- $Z_1 = X_1$ and $Z_2 = UX_1$.

(b) Use the above example to show that two Gaussian random variables can be uncorrelated, but not independent. On the other hand, show that two uncorrelated, jointly Gaussian RVs are independent.

(c) With the setup above, let $Z = VX$, where $V \in \mathbb{R}^{2\times 2}$, and $Z, X \in \mathbb{R}^2$. What is the covariance matrix Σ_Z?

(d) Use the above setup to show that $X_1 + X_2$ and $X_1 - X_2$ are independent. Give another example pair of linear combinations that are independent.

(e) Given a jointly Gaussian RV $Z \in \mathbb{R}^2$ with covariance matrix $\Sigma_Z = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix}$, how would you derive the distribution of $Z_1 | Z_2 = z$?

Hint: The following identity may be useful

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ -\frac{b}{c} & 1 \end{bmatrix} \begin{bmatrix} \left(a - \frac{b^2}{c}\right)^{-1} & 0 \\ 0 & \frac{1}{c} \end{bmatrix} \begin{bmatrix} 1 & -\frac{b}{c} \\ 0 & 1 \end{bmatrix}.$$

2 Probabilistic model of Weighted Least Squares

Let us now set up a probabilistic model from which weighted least squares arises as the natural solution.

(a) Let $X_1, X_2, \dots, X_n \in \mathbb{R}^d$ be n random vectors and $Y_1, Y_2, \dots, Y_n \in \mathbb{R}$ be one-dimensional random variables. Assume $Y_i|X_i$ are independently distributed as

$$Y_i = X_i^T w + z_i, \tag{1}$$

where $z_i \sim N(0, \sigma_i^2)$, for some fixed but unknown parameter vector $w \in \mathbb{R}^d$. What is the conditional distribution of Y_i given X_i?

(b) Derive the solution to weighted least square as a maximum likelihood estimator of the above model.

(c) Define $\tilde{Y}_i = \frac{Y_i}{\sigma_i}$ and $\tilde{X}_i = \frac{X_i}{\sigma_i}$. Suppose we still have

$$Y_i = X_i^T w + z_i, \tag{2}$$

where $z_i \sim N(0, \sigma_i^2)$.

Write out the relationship of $\tilde{X}_i$ and $\tilde{Y}_i$.

(d) Suppose $(\tilde{X}_i, \tilde{Y}_i)$ are observed for $i = 1, \dots, n$. What is the maximum likelihood estimator of w (as a function of the tuples $(\tilde{X}_i, \tilde{Y}_i)$)?

(e) You are given training data $\tilde{Y}_i = \frac{Y_i}{\sigma_i}$ and $\tilde{X}_i = \frac{X_i}{\sigma_i}$. Using part (d), derive the solution to the weighted least squares problem.