# STOR 565 Homework

1. Describe and discuss linear discriminant analysis.

2. Let $X \sim \mathcal{N}_k(\mu, \Sigma)$ and let $Y = AX + b$ where $A \in \mathbb{R}^{l \times k}$ and $b \in \mathbb{R}^l$.

   a. Find $\mathbb{E}Y$ and $\mathrm{Var}(Y)$.

   b. Argue carefully that $Y$ is multinormal and find its distribution.

   c. Fix $v \in \mathbb{R}^l$. Using the results above, find the distribution of $U = \langle v, Y \rangle$.

3. Let $\mathcal{P} = \{f_\theta : \theta > 0\}$ be the family of exponential pdfs $f_\theta(x) = \theta e^{-\theta x}$ for $x \geq 0$. Suppose that we draw $n$ samples independently from a fixed distribution $f_{\theta_0} \in \mathcal{P}$ and obtain data $x_1, \ldots, x_n \in \mathbb{R}$. The *likelihood function* for the family $\mathcal{P}$ is defined by $L(\theta) = \prod_{i=1}^n f_\theta(x_i)$. In words, $L(\theta)$ is just the joint density of the data $x_1, \ldots, x_n$ under $f_\theta$, viewed as a function of the parameter $\theta$. The *log-likelihood* is the log of the likelihood, $\ell(\theta) = \log L(\theta)$.

   a. The maximum likelihood estimate of the true parameter $\theta_0$ is defined by $\hat{\theta}_n^{\mathrm{MLE}} = \mathrm{argmax}_{\theta > 0}\, \ell(\theta)$. Use calculus to find $\hat{\theta}_n^{\mathrm{MLE}}$ in terms of the data $x_1, \ldots, x_n$.

4. Let $(X, Y)$ be a jointly distributed pair with $X \in \mathbb{R}^d$ and $Y \in \{0, 1\}$. Suppose that we have added a zeroth component to the vector $X$ that is always equal to 1, so that the augmented vector $X \in \mathbb{R}^{d+1}$. The logistic regression method for binary classification is based on the assumption that

$$\log \frac{\mathbb{P}(Y = 1 \,|\, X = x)}{\mathbb{P}(Y = 0 \,|\, X = x)} \;=\; \log \frac{\eta(x)}{1 - \eta(x)} \;=\; \langle \beta, x \rangle \tag{1}$$

for some vector $\beta \in \mathbb{R}^{d+1}$ of coefficients. In words, equation (1) says that the conditional log-odds ratio of $Y = 1$ vs. $Y = 0$ is linear in the feature vector $x$.

   a. Show, by inverting the relation (1), that

$$\eta(x) = \eta(x : \beta) \;=\; \frac{e^{\langle \beta, x \rangle}}{1 + e^{\langle \beta, x \rangle}} \;=\; \frac{1}{1 + e^{-\langle \beta, x \rangle}}$$

   Here we write $\eta(x : \beta)$ to remind ourselves that $\eta$ depends on $\beta$.

   b. Equation (1) is sometimes written in the form $\mathrm{logit}(\eta(x)) = \langle \beta, x \rangle$, where $\mathrm{logit}(u) = \log[u/(1-u)]$ for $0 < u < 1$ is the logistic (or logit) function. Sketch the logistic function.

Given a data set $D_n = (x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^{d+1} \times \{0, 1\}$ logistic regression estimates the coefficient vector $\beta$ in (1) by maximizing the conditional log likelihood function

$$\ell(\beta) \;=\; \log \prod_{i=1}^{n} \mathbb{P}_\beta(Y = y_i \,|\, X = x_i)$$

where $\mathbb{P}_\beta(Y = 1 \,|\, X = x) = \eta(x : \beta)$ and $\mathbb{P}_\beta(Y = 0 \,|\, X = x) = 1 - \eta(x : \beta)$.

c. Use the expression for $\eta(x : \beta)$ in (a) to show that the conditional log likelihood function can be written in the form

$$\ell(\beta) \;=\; \sum_{i=1}^{n} \Big[ y_i \langle \beta, x_i \rangle - \log(1 + e^{\langle \beta, x \rangle}) \Big]$$

d. Show that $\nabla \ell(\beta) = \sum_{i=1}^{n} x_i \,[y_i - \eta(x_i : \beta)]$. Hint: Evaluate the partial derivative $\partial \ell(\beta) / \partial \beta_j$ for a fixed index $j$ between 1 and $d$.

5. Let $X$ be a standard normal random variable and let $Y = X^2$.

a. Using the cdf method, find the density of $Y$.

b. Are $X$ and $Y$ independent? Why or why not?

c. What is $\text{Cov}(X, Y)$? What do these results reveal about the relationship between covariance and independence?

6. Let $X \geq 0$ be a random variable with $\mathbb{E}X = 10$ and $\mathbb{E}X^2 = 140$.

a. Find an upper bound on $\mathbb{P}(X > 14)$ involving $\mathbb{E}X$ using Markov's inequality.

b. Modify the proof of Markov's inequality to find an upper bound on $\mathbb{P}(X > 14)$ involving $\mathbb{E}X^2$.

c. Compare the results in (a) and (b) above to what you find from Chebyshev's inequality.

7. Let $X$ be a random variable with $\text{Var}(X) = 3$. Use Chebyshev's inequality to find upper bounds on $\mathbb{P}(|X - \mathbb{E}X| > 1)$ and $\mathbb{P}(|X - \mathbb{E}X| > 2)$. Comment on the potential usefulness of these bounds.

8. Recall that the moment generating function of a random variable $X$ is defined by $M_X(s) = \mathbb{E}e^{sX}$ for all $s$ such that the expectation is finite. Find the moment generating function (MGF) of the following distributions.

a. Poisson($\lambda$)

b. $\mathcal{N}(0, 1)$

9. Find the gradient and Hessian of the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x) = x_1^2 x_2 + 3x_1 - 5x_2 + 1$$

10. State and prove Markov's probability inequality.

11. Let $X$ and $Y$ be random variables with moment generating functions $M_X(s)$ and $M_Y(s)$, respectively. Show that $S = X+Y$ has moment generating function $M_S(s) = M_X(s)\, M_Y(s)$.