CS 534: Machine Learning
Homework 0

(Due Aug 27th at 11:59 PM on Gradescope)

1. Have you read through the course website, noted the important dates, and the class policies? (yes or no).

2. Which of the following courses have you taken?
   (i) Have you taken any course on Probability/Statistics? If yes, please write down the institution name, course department, course name, and your course grade.
   (ii) Have you taken any course on Linear Algebra? If yes, please write down the institution name, course department, course name, and your course grade.
   (iii) Have you taken any course on Optimization? If yes, please write down the institution name, course department, course name, and your course grade.
   (iv) Have you taken any courses on Data Mining/Pattern Recognition/Machine Learning? If yes, please write down the institution name, course department(s), course name(s), and your course grade(s).

3. An urn has 3 red balls, 4 blue balls, and 5 green balls. Alice draws a ball from the urn, and then Bob draws a ball. What is the probability that Bob got a green ball?

4. Consider a 3-valued random variable X such that $P(X = 1) = 0.35$, $P(X = 0) = 0.45$ and $P(X = -1) = 0.2$. Assume you have access to a program A that generates a number in $[0, 1]$ uniformly at random. Describe how you can use A to draw random samples of X.

5. In your favorite programming language, implement the program above to draw 100 random samples of X.

6. Consider the standard basis for $\mathbb{R}^n$: $\mathbf{e}_1 = [1, 0, 0, \cdots, 0], \mathbf{e}_2 = [0, 1, 0, \cdots, 0], \cdots, \mathbf{e}_p = [0, 0, 0, \cdots, 1]$. Recall that the inner-product of two vectors $\mathbf{w}_1 = [\alpha_1, \alpha_2, \cdots, \alpha_n], \mathbf{w}_2 = [\beta_1, \beta_2, \cdots, \beta_n] \in \mathbb{R}^n$ is given by:

   $$< \mathbf{w}_1, \mathbf{w}_2 >= \sum_{i=1}^{n} \alpha_i \beta_i.$$

   Let $f : \mathbb{R}^n \to \mathbb{R}$ be a linear map. Show there exists a vector $\mathbf{x} \in \mathbb{R}^n$ such that:

   $$f(\mathbf{w}) =< \mathbf{x}, \mathbf{w} >, \text{ for any } \mathbf{w} \in \mathbb{R}^n.$$

   (Hint) A linear map has the following properties that you may find useful:

   (i) $f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$, for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$
   (ii) $f(\kappa \mathbf{x}) = \kappa f(\mathbf{x})$, for $\mathbf{x} \in \mathbb{R}^n, \kappa$ scalar.

7. Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ and $\mathbf{y} \in \mathbb{R}^n$ be given.

   (i) Find the optimal vector $\mathbf{w}^* \in \mathbb{R}^p$ which solves the following problem:

   $$\min_{\mathbf{w} \in \mathbb{R}^p} \frac{1}{2} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2$$

   (Hint) Consult the *Matrix Cookbook* if you want to look up expressions for derivatives in matrix/vector form.

(ii) Does your solution above work if $\mathbf{X}$ is not full rank? If not, name one way to compute $\mathbf{w}^*$.

(iii) What is the optimal solution to the following problem?

$$\min_{w \in \mathbb{R}^p} \frac{1}{2}\|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2 + \frac{c}{2}\|\mathbf{w}\|^2, \text{ where } c > 0 \text{ is a constant}$$

8. What is the probability density function $p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ of a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$? Please provide an expression in terms of $\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}$, and clearly define any special function you use in the expression.

9. Let $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$ be the precision or inverse covariance matrix. What is expression of the probability density function $p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Theta}^{-1})$ of a multivariate Gaussian distribution in terms of the mean $\boldsymbol{\mu}$ and precision matrix $\boldsymbol{\Theta}$?

10. For a bivariate Gaussian with mean $\boldsymbol{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}$ and covariance $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$, where ρ is the correlation and $\sigma_X, \sigma_Y > 0$, what is the:

    (i) Marginal distribution of Y, $P(Y)$?

    (ii) Conditional distribution of X, $P(X|Y)$?