CSE 847: Machine Learning— Midterm Sample Questions

1. True or False

   (a) The singular value decomposition of a matrix $A \in \mathbb{R}^{m \times n}$ is unique.

   (b) If $A \in \mathbb{R}^{n \times n}$ is positive definite, then all its real eigenvalues must be positive.

   (c) If $B \in \mathbb{R}^{m \times m}$ and $C \in \mathbb{R}^{n \times n}$ are both nonsingular, then $\text{rank}(A) = \text{rank}(BAC)$, for any matrix $A \in \mathbb{R}^{m \times n}$.

   (d) In a least-squares linear regression problem, adding an ℓ_2 regularization penalty always decreases the expected ℓ_2 error of the solution w^* on unseen test data.

   (e) If K is valid kernel, then $K + \alpha I$ must be a valid kernel, where α is a scalar.

2. You have an old box containing 2 apples and 8 oranges and a new box containing 4 apples and 6 oranges. You select a box at random with equal probability, select an item out of that box at random with equal probability, and find it is an apple. Using Bayes' theorem to find the probability that the apple came from the old box?

3. Describe in details the three distinct approaches (probabilistic generative models, probabilistic discriminative models, and discriminate function) for modeling.

4. Consider two data sets $\mathcal{D}_1$ and $\mathcal{D}_2$, each consisting of scalar measurements $x_i, i = 1, ..., N_1$ for $\mathcal{D}_1$ and $x_j, j = 1, ..., N_2$ for $\mathcal{D}_2$. Assume that each set of measurements comes from a Gaussian distribution. The two Gaussian distributions share a common variance σ^2 and the mean μ. μ and σ^2 are both assumed unknown.

   (a) Define the log-likelihood for this problem.

   (b) Derive the maximum likelihood estimators for the unknown parameters.

5. Given a data matrix $A \in \mathbb{R}^{n \times d}$ and $y \in \mathbb{R}^d$, the ℓ_2-norm regularized ℓ_2 loss function for the least squares regression is given by:

$$\min_x \frac{1}{2}\|Ax - y\|_2^2 + \frac{\lambda}{2}\|x\|_2^2 \tag{P}$$

   (a) Suppose we have a typo and write the formulation in: $\min_x \frac{1}{2}\|Ax - y\|_2^2 + \frac{\lambda}{2}\|y\|_2^2$. What is gradient of this wrong objective with respect to x? What is the effect of the regularization?

   (b) Suppose we use the correct formulation in (P) but choose $\lambda < 0$. Describe briefly what is the effect of the regularization and why.