**Name:** **Student ID:**

# CS189: Introduction to Machine Learning

## Homework 2

Due: February 17, 2015 at 11:59pm

**Problem 1.** A target is made of 3 concentric circles of radii $1/\sqrt{3}$, 1 and $\sqrt{3}$ feet. Shots within the inner circle are given 4 points, shots within the next ring are given 3 points, and shots within the third ring are given 2 points. Shots outside the target are given 0 points.

Let $X$ be the distance of the hit from the center (in feet), and let the p.d.f of $X$ be

$$f(x) = \begin{cases} \frac{2}{\pi(1+x^2)} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

What is the expected value of the score of a single shot?

**Solution:**

**Problem 2.** Assume that the random variable $X$ has the exponential distribution

$$f(x|\theta) = \theta e^{-\theta x} \qquad x > 0, \theta > 0$$

where $\theta$ is the parameter of the distribution. Use the method of maximum likelihood to estimate $\theta$ if 5 observations of $X$ are $x_1 = 0.9$, $x_2 = 1.7$, $x_3 = 0.4$, $x_4 = 0.3$, and $x_5 = 2.4$, generated i.i.d.

**Solution:**

**Problem 3.** The polynomial kernel is defined to be

$$k(\mathbf{x}, \mathbf{y}) = (\mathbf{x^T y + c})^{\mathbf{d}}$$

where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathbf{n}}$, and $c \geq 0$. When we take $d = 2$, this kernel is called the quadratic kernel.

(a) Find the feature mapping $\Phi(\mathbf{z})$ that corresponds to the quadratic kernel.

(b) How do we find the optimal value of $d$ for a given dataset?

**Solution:**

**Def**: Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. We say that $A$ is positive definite if $\forall x \in \mathbb{R}^n,\ x^\top A x > 0$. Similarly, we say that $A$ is positive semidefinite if $\forall x \in \mathbb{R}^n,\ x^\top A x \geq 0$.

**Problem 4.** Let $x = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top \in \mathbb{R}^n$, and let $A \in \mathbb{R}^{n \times n}$ be the square matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

(a) Give an explicit formula for $x^\top A x$. Write your answer as a sum involving the elements of $A$ and $x$.
**Solution:**

(b) Show that if $A$ is positive definite, then the entries on the diagonal of $A$ are positive (that is, $a_{ii} > 0$ for all $1 \leq i \leq n$).
**Solution:**

**Problem 5.** Let $B$ be a positive semidefinite matrix. Show that $B + \gamma I$ is positive definite for any $\gamma > 0$.

**Solution:**

**Problem 6.** Suppose we have a classification problem with classes labeled $1, \ldots, c$ and an additional doubt category labeled as $c+1$. Let the loss function be the following:

$$\ell(f(x) = i, y = j) = \begin{cases} 0 & \text{if } i = j \quad i, j \in \{1, \ldots, c\} \\ \lambda_r & \text{if } i = c+1 \\ \lambda_s & \text{otherwise} \end{cases}$$

where $\lambda_r$ is the loss incurred for choosing doubt and $\lambda_s$ is the loss incurred for making a misclassification. Note that $\lambda_r \geq 0$ and $\lambda_s \geq 0$.

(a) Show that the minimum risk is obtained if we follow this policy: (1) choose class $i$ if $P(\omega_i|x) \geq P(\omega_j|x)$ for all $j$ and $P(\omega_i|x) \geq 1 - \lambda_r/\lambda_s$, and (2) choose doubt otherwise.
**Solution:**

(b) What happens if $\lambda_r = 0$? What happens if $\lambda_r > \lambda_s$?
**Solution:**

**Problem 7.** Let $p(x|\omega_i) \sim \mathcal{N}(\mu_i, \sigma^2)$ for a two-category one-dimensional classification problem with $P(\omega_1) = P(\omega_2) = 1/2$.

(a) Show that the minimum probability of error is

$$P_e = \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-(1/2)u^2} du$$

where $a = |\mu_2 - \mu_1|/2\sigma$.
**Solution:**

(b) Use the inequality

$$\frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-(1/2)u^2} du \leq \frac{1}{\sqrt{2\pi}a} e^{-(1/2)a^2} du$$

to show that $P_e$ goes to zero as $a = |\mu_2 - \mu_1|/\sigma$ goes to infinity.
**Solution:**

**Problem 8.** Recall that the probability mass function of a Poisson random variable is

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!} \qquad x \in \{0, 1, \ldots, \infty\}$$

You are given two *equally likely* classes of Poisson data with parameters $\lambda_1 = 10$ and $\lambda_2 = 15$. This means that $x|\omega_1 \sim \text{Poisson}(\lambda_1)$ and $x|\omega_2 \sim \text{Poisson}(\lambda_2)$.

(a) Given the class conditionals, $x|\omega_1$ and $x|\omega_2$, find $P(\omega_1|x)$ in terms of $\lambda_1$, $\lambda_2$, $P(\omega_1)$, and $P(\omega_2)$. What type of function is the posterior?
**Solution:**

(b) Find the optimal rule (decision boundary) for allocating an observation $x$ to a particular class. Calculate the probability of correct classification for each class. Calculate the total error rate for this choice of decision boundary.
**Solution:**

(c) Suppose instead of one, we can obtain two independent measurements $x_1$ and $x_2$ for the object to be classified. How do the allocation rules and error rates change? Calculate the revised probability of correct classification for each class. Calculate the new total error in this case.

*Hint:* Always keep in mind that the Poisson distribution is defined for nonnegative integral values. Moreover, you can't be sure how much error you accumulate by erring on either side unless you explicitly calculate it.

**Solution:**

**Problem 9 (Optional: Extra for Experts) .** Let $X_1, X_2, ..., X_n$ be a sequence of points chosen independently and uniformly from within a 2-dimensional unit ball $B = \{x \in \mathbb{R}^2 : {x_1}^2 + {x_2}^2 \leq 1\}$. A set of points $X_1, X_2, ...X_n$ lie in a hemisphere if there is a line passing through the origin for which all $n$ points lie on a particular side of the hemisphere. Define the event:

$$A_n = \{X_1, X_2...X_n \text{ lie in a hemisphere}\}$$

Compute $\Pr\{A_n\}$. (There are multiple ways of doing this. Some are simpler than others)

Credit and Thanks to Professor Thomas Courtade for writing this question

**Solution:**