CS 274A Homework 3


Due Date: submit hardcopy at the start of class on Monday Feb 5th

Instructions and Guidelines for Homeworks

- Please answer all of the questions and submit a hardcopy of your written solutions (either hand-written or typed are fine as long as the writing is legible). Clearly mark your name on the first page.

- All problems are worth 10 points unless otherwise stated. All homeworks will get equal weight in computation of the final grade for the class.

- The homeworks are intended to help you work through the concepts we discuss in class in more detail. It is important that you try to solve the problems yourself. The homework problems are important to help you better learn and reinforce the material from class. If you don’t do the homeworks you will likely have difficulty in the exams later in the quarter.

- If you can’t solve a problem, you can discuss it verbally with another student. However, please note that before you submit your homework solutions you are not allowed to view (or show to any other student) any written material directly related to the homeworks, including other students’ solutions or drafts of solutions, solutions from previous versions of this class, and so forth. The work you hand in should be your own original work.

- You are allowed to use reference materials in your solutions, such as class notes, textbooks, other reference material (e.g., from the Web), or solutions to other problems in the homework. It is strongly recommended that you first try to solve the problem yourself, without resorting to looking up solutions elsewhere. If you base your solution on material that we did not discuss in class, or is not in the class notes, then you need to clearly provide a reference, e.g., “based on material in Section 2.2 in .....”

- In problems that ask for a proof you should submit a complete mathematical proof (i.e., each line must follow logically from the preceding one, without “hand-waving”). Be as clear as possible in explaining your notation and in stating your reasoning as you go from line to line.

- If you wish to use LaTeX to write up your solutions you may find it useful to use the .tex file for this homework that is posted on the Web page.
Problem 1: (Multivariate Gaussians)

Let $\mathbf{x} = (x_1, x_2, \ldots, x_d)$ be a real-valued $d$-dimensional vector with real-valued components. Assume that $p(\mathbf{x})$ is Gaussian with mean $\mu$ and covariance matrix $\Sigma$. Prove that if $\Sigma$ is a diagonal matrix that this implies that the variables $X_1, X_2, \ldots, X_d$ are independent. (For a definition of the multivariate Gaussian model see Section 3.2 of Note Set 2 on the class Website).

Problem 2: (Maximum Likelihood for a Gaussian with a Diagonal Covariance Matrix)

Assume that $\mathbf{x}$ is $d$-dimensional and real-valued and has a Gaussian distribution with a mean vector $\mu$ and covariance matrix $\Sigma$. Further assume that the covariance matrix $\Sigma$ is diagonal with entries $\sigma_1^2, \sigma_2^2, \ldots, \sigma_d^2$ on the diagonals. Derive the maximum likelihood estimates for $\mu$ and for the variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_d^2$, given observed data $D = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, where it is assumed that the $\mathbf{x}_i$ are conditionally independent given $\mu$ and $\Sigma$. State precisely any assumptions you make in your solution.

Problem 3: (Maximum Likelihood for Naive Bayes)

Consider a naive Bayes classifier with a random variable $C$ taking values $c \in \{1, \ldots, K\}$ (referred to as the “class variable”) and a $d$-dimensional “feature” random variable $X = (X_1, \ldots, X_d)$ taking values $\mathbf{x} = (x_1, \ldots, x_d)$. Here we will assume that each component of $X$ is a binary random variable (i.e., the $x_i$’s take values 0 or 1, although in general the $x_i$’s could be $M$-ary, real-valued, etc. ). In the naive Bayes classification model we assume that the feature random variables $X_1, \ldots, X_d$ are all conditionally independent of each other given $C$. The naive Bayes model is a simple classification model, where the goal in classification is to compute $P(c = k|\mathbf{x})$ given a feature vector $\mathbf{x}$ and where the value of the class $C$ is known.

Consider that we have data $D = \{(\mathbf{x}_i, c_i)\}, 1 \leq i \leq N$. Here $\mathbf{x}_i = (x_{i1}, \ldots, x_{ij}, \ldots, x_{id})$ is the $i$th observed feature vector where $x_{ij}$ is the $j$th component (or $j$th feature value) for the $i$th feature vector, and $c_i$ is the observed class label for $\mathbf{x}_i$. Let $\theta_j^{(k)} = P(x_j = \mid c = k), 1 \leq j \leq d, 1 \leq k \leq K$ and let $\beta_k = P(c = k).

1. Define the likelihood $L(\theta, \beta)$ for this problem where $\theta$ represents the set of all of the $\theta_j^{(k)}$’s and $\beta$ the set of $\beta_k$’s.

2. Derive the maximum likelihood estimates for $\theta_j^{(k)}$ and $\beta_k$.

Problem 4: (Bayesian Estimation of a Gaussian Model)

For the case of the mean $\mu$ of a Gaussian model, with known variance $\sigma^2$, and with a Gaussian prior on $\mu$ that has mean $\mu_0$ and variance $\sigma^2$, we discussed in class the fact that the posterior density for $\mu$, given $n$ IID
Homework 3: CS 274A, Probabilistic Learning: Winter 2018

observations \{x_1, \ldots, x_n\} is also Gaussian with parameters \(\mu_n\) and \(\sigma_n^2\). Prove that the following expressions the mean of this posterior and the variance of this posterior are correct, i.e.,

\[
\mu_n = \gamma \hat{\mu}_{ML} + (1 - \gamma) \mu_0
\]

where

\[
\gamma = \frac{ns^2}{ns^2 + \sigma^2},
\]

and

\[
\frac{1}{\sigma_n^2} = \frac{n}{\sigma^2} + \frac{1}{s^2}.
\]

**Problem 5: (Bayesian Estimation for the Multinomial Model)**

Consider a data set \(D = \{x_1, \ldots, x_N\}\), with \(x_i \in \{1, \ldots, M\}\) where the \(x_i\)'s are independent draws from the discrete distribution with parameters \(\theta_k = P(x_i = k), 1 \leq k \leq M, \) and \(\sum_{k=1}^M \theta_k = 1\) (i.e., we have a multinomial model and likelihood for the \(x_i\)'s). Assume that we have a Dirichlet prior for the parameters \(\theta\), where the prior has parameters \(\alpha_1, \ldots, \alpha_M\) and \(\alpha_k > 0, 1 \leq k \leq M\).

1. Prove that the posterior distribution on \(\theta_1, \ldots, \theta_K\) is also Dirichlet.

2. Derive an expression for the maximum a posteriori (MAP) estimate for \(\theta_k, 1 \leq k \leq M\). Your solution should be from first principles (i.e. using basic calculus working from your solution for of \(P(\theta|D)\) from part 1).

**Problem 6: (Bayesian Estimation for the Gamma-Exponential)**

Consider a data set \(D = \{x_1, \ldots, x_N\}\), where \(x_i\)'s are real-valued and \(x_i > 0, 1 \leq i \leq N\), and where the \(x_i\)'s are independent draws from the exponential density \(p(x|\theta) = \theta e^{-\theta x}\).

Define a Gamma prior for \(\theta\) in the form \(p(\theta|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta \theta}\), where \(\alpha > 0\) and \(\beta > 0\) are the parameters of the Gamma prior and \(\Gamma(.)\) is the special Gamma function. Note that when \(\alpha = 1\) we get an exponential density (on \(\theta\), with parameter \(\beta\)) as a special case.

1. Derive an expression for the posterior density on \(\theta\) and state what type of density this.

2. What are the posterior mode and posterior mean for this model? For this problem you can state the results, and briefly explain your reasoning, you don’t need to derive the results from first principles unless you want to.
Problem 7: (Bayesian Estimation for the Uniform Model)

Let $X$ be uniformly distributed with lower limit $0$ and an unknown upper limit $\theta$, where $\theta > 0$, i.e.,

$$p(x; \theta) = \frac{1}{\theta}$$

for $0 \leq x \leq \theta$ and $p(x) = 0$ otherwise. Let $D = \{x_1, \ldots, x_N\}$ be an observed data set of $N$ random sample from $p(x; \theta)$. This is the same setup as in Homework 2 but here we will look at a Bayesian approach.

1. Let $P(\theta) = \frac{1}{(\alpha_2 - \alpha_1)}$ be a prior on $\theta$, where $\alpha_2 > \alpha_1 > 0$. Assume that $0 \leq x_i \leq \alpha_2, 1 \leq i \leq N$. Derive an expression for the posterior density $P(\theta|D)$ with this prior.

2. Let $P(\theta) = \lambda e^{-\lambda \theta}$ be a different prior on $\theta$, where $\lambda > 0$. Derive an expression for the posterior density $P(\theta|D)$ with this prior.

For both parts of the problem above your solution should explain how the normalizing constant for $P(\theta|D)$ is computed.

(Optional, will not be graded): derive $\hat{\theta}_{MAP}$ and $\hat{\theta}_{MPE}$ for both cases above.

Problem 8: (Computer Simulation for the Beta-Binomial Model)

Write code (e.g., in Matlab or Python or R) to do the following. Submit a hardcopy of the figures (with 9 subplots in each figure) as requested below. Also comment briefly on what can be learned from these plots (a few sentences are fine).

Consider the binomial likelihood model with a beta prior as we discussed in class. Let the number of successes in the data be $r$ out of $n$ trials. Consider the following 3 data sets:

1. $r = 4, n = 5$;
2. $r = 40, n = 50$;
3. $r = 400, n = 500$;

For each of the above data sets you are to generate a single figure, where each figure contains 9 different plots arranged in 3 rows and 3 columns. (e.g., in Matlab you can use the function subplot.m to generate multiple plots on a single figure in this fashion).

The plots in rows 1, 2, and 3 should correspond to “prior strengths” of $\alpha + \beta = 1, 10, 100$ respectively. The plots in columns 1, 2, and 3 should correspond to prior means of $0.2, 0.5, \text{ and } 0.8$ respectively. For each individual plot, plot the prior $p(\theta)$ and the posterior $p(\theta|D)$. The prior should be plotted using a dotted line, the posterior using a solid line. (If you have access to a color printer you can use red and green colors (for the prior and posterior respectively) for each line). In summary, you will be generating 3 separate figures, where each contains $3 \times 3 = 9$ plots. In your hardcopy submission please put each of the 3 figures on a separate page.