Homework Set Four
ECE 271A
Department of Computer and Electrical Engineering
University of California, San Diego

This HW set contains several problems. Only the problem labeled **Quiz** must be handed in and will be graded. The remaining problems are for practice. You should not submit them for grade. By submitting your Quiz solution, you agree to comply with the following.

1. The Quiz is treated as a **take-home test** and is an **INDIVIDUAL** effort. **NO collaboration is allowed**. The submitted work must be yours and must be original.
2. The work that you turn-in is your own, using the resources that are available to **all** students in the class.
3. You can use the help of **GENERAL** resources on programming, such as MATLAB tutorials, or related activities.
4. You are not allowed to consult or use resources provided by tutors, previous students in the class, or any websites that provide solutions or help in solving assignments and exams.
5. You will not upload your solutions or any other course materials to any web-sites or in some other way distribute them outside the class.
6. 0 points will be assigned if your work seems to violate these rules and, if recurrent, the incident(s) will be reported to the Academic Integrity Office.

1. Bayesian regression: in last week's problem set we showed that various forms of linear regression by the method of least squares are really just particular cases of ML estimation under the model

$$\mathbf{z} = \mathbf{\Phi}\theta + \epsilon$$

where $\mathbf{z} = (z_1, \ldots, z_n)^T$, $\theta = (\theta_1, \ldots, \theta_k)^T$

$$\mathbf{\Phi} = \begin{bmatrix} 1 & \ldots & x_1^K \\ \vdots & & \vdots \\ 1 & \ldots & x_n^K \end{bmatrix}$$

and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^T$ is a normal random process $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. It seems only natural to consider the Bayesian extension of this model, an extension that has been the subject of some recent research under the denomination of *Gaussian processes*. For this, we simply extend the model considering a Gaussian prior

$$P_\theta(\theta) = \mathcal{G}(\theta, \mathbf{0}, \mathbf{\Gamma}).$$

a) Given a training set $\mathcal{D} = \{(\mathcal{D}_x, \mathcal{D}_z)\} = \{(x_1, z_1), \ldots, (x_n, z_n)\}$, compute the posterior distribution

$$P_{\theta|\mathbf{T}}(\theta|\mathcal{D})$$

and the predictive distribution

$$P_{z|\mathbf{T}}(z|\mathcal{D}).$$

b) Consider the MAP estimate

$$\theta_{MAP} = \arg\max_{\theta} P_{\theta|\mathbf{T}}(\theta|\mathcal{D}).$$

How does it differ from the weighted least squares estimate? What is the role of the terms that were not present in the latter? Is there any advantage in setting them to anything other than zero?

c) Consider the case in which prior covariance $\mathbf{\Gamma}$ is a diagonal matrix, not necessarily the identity. Suppose that you are told that K, i.e. the number of parameters in θ or the degree of the polynomial $\phi(x)^T\theta$, is somewhere between 1 and 25. How would you set up $\mathbf{\Gamma}$ and why? Discuss the implications of your selection on the bias and variance of your MAP solution

$$z_{MAP} = \mathbf{\Phi}(x)\theta_{MAP}.$$

2. In this problem we explore the exponential family and conjugate priors. The exponential family is the family of densities of the form

$$P_{\mathbf{X}|\theta}(\mathbf{x}|\theta) = f(\mathbf{x})g(\theta)e^{\phi(\theta)^T u(\mathbf{x})}$$

with

$$[g(\theta)]^{-1} = \int f(\mathbf{x})e^{\phi(\theta)^T u(\mathbf{x})}d\mathbf{x}.$$

a) Show that, for a density in this family, the likelihood of a sequence $\mathcal{D} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is

$$P_{\mathbf{T}|\theta}(\mathcal{D}|\theta) \propto \prod_{i=1}^{n} f(\mathbf{x}_i) \exp\left\{\phi(\theta)^T \sum_{i=1}^{n} u(\mathbf{x}_i)\right\}.$$

What is the normalization constant?

b) It has been shown that, apart from certain irregular cases, the exponential family is the only family of distributions for which there is a conjugate prior. Show that

$$P_\theta(\theta) = \frac{g(\theta)^\eta e^{\phi(\theta)^T \nu}}{\int g(\theta)^\eta e^{\phi(\theta)^T \nu} d\theta}$$

is a conjugate prior for the exponential family and compute the posterior distribution $P_{\theta|\mathbf{T}}(\theta|\mathcal{D})$. Denoting $\mathbf{s} = \sum_{i=1}^{n} u(\mathbf{x}_i)$ as *the sufficient statistic*, compare the posterior with prior density. What is the result of "propagating" the prior through the likelihood function?

c) Consider table 1. For each row **i)** show that the likelihood function on the right column belongs to the exponential family, **ii)** show that the prior on the left column is a conjugate prior for the likelihood function on the right column, **iii)** compute the posterior $P_{\theta|\mathbf{T}}(\theta|\mathcal{D})$, and **iv)** interpret the meaning of the sufficient statistic and the "propagation" discussed in **b)**.

d) Repeat the steps of **c)** for the distributions of problem **4.**, i.e. the **multinomial** as the likelihood function and the **Dirichlet** as the prior.

3. (Quiz) Finish up the Quiz of assignment **3**.

Likelihood	$P_{\mathbf{T}\vert\theta}(\mathcal{D}\vert\theta)$	**Prior**	$P_\theta(\theta)$
Bernoulli	$\prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}$	**Beta**	$P_\theta(\theta;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$
Poisson	$\prod_{i=1}^n \frac{e^{-\theta}\theta^{x_i}}{x_i!}$	**Gamma**	$P_\theta(\theta;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}$
Exponential	$\prod_{i=1}^n \theta e^{-\theta x_i}$	**Gamma**	$P_\theta(\theta;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}$
Normal ($\theta = 1/\sigma^2$)	$\prod_{i=1}^n \sqrt{\frac{\theta}{2\pi}}\exp\{-\frac{\theta}{2}(x_i-\mu)^2\}$	**Gamma**	$P_\theta(\theta;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}$

Table 1: In the case of the normal distribution, μ is assumed known, the parameter is the precision $\theta = 1/\sigma^2$.