CS181 Practice Questions: Probability and Linear Regression Basics

1. **Mean of Gaussian (Bishop, 1.8, part 1)**

   By using a change of variables, verify that the univariate Gaussian distribution satisfies

$$
\begin{aligned}
\mathbb{E}[x] &= \int (2\pi\sigma^2)^{-1/2} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\} x \, dx \\
&= \mu.
\end{aligned}
$$

2. **Mode of Gaussian (Bishop, 1.9)**

Show that the mode (i.e. the maximum) of the Gaussian distribution

$$\mathcal{N}(x|\mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\{-(2\sigma^2)^{-1}(x-\mu)^2\}$$

is given by μ.

3. **Gaussian MLE**

Suppose we have N iid values $x_n \sim \mathcal{N}(\mu, \sigma^2)$, where $n = 1, \ldots, N$.

(a) Write down the likelihood function.

(b) Write down the log-likelihood function.

(c) Find the maximum likelihood estimator for μ_{ML}.

(d) Find the maximum likelihood estimator for σ^2_{ML}.

(e) Show that the μ_{ML} is unbiased.

(f) Show that the σ^2_{ML} is biased.

(g) Give an unbiased estimator for the variance parameter.

4. **MLE Estimate of the Bias Term (Bishop (3.19))**

Let $\mathbf{\Phi}$ be our $N \times J$ design matrix, $\boldsymbol{t}$ our vector of N target values, $\boldsymbol{w}$ our vector of J parameters, and w_0 our bias parameter. As Bishop notes in (3.18), the sum-of-squares error function of $\boldsymbol{w}$ and w_0 can be written as follows

$$E(\boldsymbol{w}, w_0) = \frac{1}{2} \sum_{n=1}^{N} \left(t_n - w_0 - \sum_{j=1}^{J-1} w_j \cdot \phi_j(x_n) \right)^2 .$$

Show that the value of w_0 that minimizes E is

$$\begin{aligned} w_{0_{MLE}} &= \frac{1}{N} \sum_{n=1}^{N} t_n - \sum_{j=1}^{J-1} w_j \cdot \left(\frac{1}{N} \sum_{n=1}^{N} \phi_j(x_n) \right) \\ &= \bar{t} - \sum_{j=1}^{J-1} w_j \cdot \overline{\phi_j(x)} \qquad \text{[compare Bishop (3.19)]} \end{aligned}$$

5. **Simple Linear Regression (Bishop, 1.1)**

Consider the sum-of-squares error function given by:

$$E(\boldsymbol{w}) = \frac{1}{2} \sum_{n=1}^{N} \{y(x_n, \boldsymbol{w}) - t_n\}^2,$$

in which the function $y(x, \boldsymbol{w})$ is given by the polynomial

$$y(x, \boldsymbol{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j.$$

Show that the coefficients $\boldsymbol{w} = \{w_i\}$ that minimize this error function are given by the solution to the following set of linear equations

$$\sum_{j=0}^{M} A_{ij} w_j = T_i$$

where

$$A_{ij} = \sum_{n=1}^{N} (x_n)^{i+j},$$
$$T_i = \sum_{n=1}^{N} (x_n)^i t_n.$$

Here a suffix i or j denotes the index of a component, where as $(x)^i$ denotes x reaised to the power of i.

6. **Multivariate Regression (Adapted from Stanford CS229)**

Suppose we have $\mathbf{\Phi} \in \mathbb{R}^{N \times J}$ as our design matrix, but that instead of predicting scalar values t_n, we'd like to use least squares regression to predict vector-valued targets $\boldsymbol{t}_n \in \mathbb{R}^M$ for each row $\phi(\boldsymbol{x}_n)$ in $\mathbf{\Phi}$. To do this, we can introduce a parameter matrix $\boldsymbol{W} \in \mathbb{R}^{N \times M}$ and attempt to minimize the following sum-of-squares error function:

$$
\begin{aligned}
E(\boldsymbol{W}) &= \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{M} \left((\boldsymbol{W}^\top \phi(\boldsymbol{x}_n))_m - t_{nm} \right)^2 \\
&= \frac{1}{2} \sum_{n=1}^{N} (\boldsymbol{W}^\top \phi(\boldsymbol{x}_n) - \boldsymbol{t}_n)^\top (\boldsymbol{W}^\top \phi(\boldsymbol{x}_n) - \boldsymbol{t}_n) \\
&= \frac{1}{2} \mathrm{tr}((\mathbf{\Phi}\boldsymbol{W} - \boldsymbol{T})^\top (\mathbf{\Phi}\boldsymbol{W} - \boldsymbol{T})),
\end{aligned}
$$

where $\boldsymbol{T} = \begin{bmatrix} — & \boldsymbol{t}_1^\top & — \\ & \vdots & \\ — & \boldsymbol{t}_N^\top & — \end{bmatrix}$.

(a) Rewrite $E(\boldsymbol{W})$ in terms of the traces of 3 matrices.

(b) Derive the gradient of your answer to part (a) with respect to $\boldsymbol{W}$.

(c) Derive $\boldsymbol{W}_{MLE}$ by setting the gradient of your answer to part (b) to 0.

Hint: in answering the above questions, it may be useful to keep in mind the following facts:

1. $\mathrm{tr}(\boldsymbol{A} + \boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{B})$
2. $\mathrm{tr}(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{A}^\top)$
3. $\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{B}\boldsymbol{A})$
4. $\frac{\partial}{\partial \boldsymbol{A}} \mathrm{tr}(\boldsymbol{A}\boldsymbol{B}) = \boldsymbol{B}^\top$
5. $\frac{\partial}{\partial \boldsymbol{A}} \mathrm{tr}(\boldsymbol{A}^\top \boldsymbol{B}) = \boldsymbol{B}$