CS725: Assignment 2

25 Marks, Due on September 8^{th} in the class

1. You are told that the pdf for a continuous valued random variable is non-zero only for $x \geq 0$ and that its mean is μ. What will be the form of the distribution that maximizes its entropy, given these conditions?

   (**5 Marks**)

2. Consider a linear regression model:

   $$Y = \mathbf{w}^T \phi(\mathbf{X}) + \epsilon$$

   where $\phi(\mathbf{X}) \in \Re^n$ and $\epsilon \sim \mathcal{N}(0, \sigma^2)$. In actuality, let only the first $n_1 \leq n$ basis functions be important, *i.e.*, let $\mathbf{w} = \left[\mathbf{w}[1:n_1]^T,\ \mathbf{0}^T\right]^T$ where, $\mathbf{w}[1:k]$ is a vector with the first k elements from $\mathbf{w}$. Similarly, let $\phi(\mathbf{X})[1:k]$ denote the vector of the first k components of $\phi(\mathbf{X})$. Let

   $$\widehat{\mathbf{w}}_{MLE} = \underset{\mathbf{w} \in \Re^n}{\operatorname{argmax}} \ \mathrm{LL}(y_1, \ldots, y_n \ | \phi(\mathbf{X}_1), \ldots, \phi(\mathbf{X}_m), \mathbf{w})$$

   And for any $k \leq n$,

   $$\widehat{\mathbf{w}}^k_{MLE} = \underset{\mathbf{w} \in \Re^k}{\operatorname{argmax}} \ \mathrm{LL}(y_1, \ldots, y_n \ | \phi(\mathbf{X}_1)[1:k], \ldots, \phi(\mathbf{X}_m)[1:k], \mathbf{w})$$

   Which of the following statement(s) is/are true? Prove.

   (a) A consequence of including irrelevant basis functions (in this case, $\phi_{n_1+1} \ldots \phi_n$) in a regression model is that the predictions have smaller variances but they are biased. That is,

   $$E\left[\phi(\mathbf{X})^T \widehat{\mathbf{w}}_{MLE}\right] \neq \phi(\mathbf{X})[1:n_1]^T \mathbf{w}[1:n_1]$$

   and that

   $$Var\left[\phi(\mathbf{X})^T \widehat{\mathbf{w}}_{MLE}\right] \leq Var\left[\phi(\mathbf{X})[1:n_1]^T \widehat{\mathbf{w}}^{n_1}_{MLE}\right]$$

(b) A consequence of including irrelevant basis functions (in this case, $\phi_{n_1+1} \ldots \phi_n$) in a regression model is that the predictions have larger variances though they are unbiased. That is,

$$E\left[\phi(\mathbf{X})^T \widehat{\mathbf{w}}_{MLE}\right] = \phi(\mathbf{X})[1:n_1]^T \mathbf{w}[1:n_1]$$

and that

$$Var\left[\phi(\mathbf{X})^T \widehat{\mathbf{w}}_{MLE}\right] \geq Var\left[\phi(\mathbf{X})[1:n_1]^T \widehat{\mathbf{w}}^{n_1}_{MLE}\right]$$

(c) A consequence of excluding relevant basis functions in a linear model is that the predictions are biased, though they have smaller variances. That is, for all integers $n_2 \in (0, n_1)$,

$$E\left[\phi(\mathbf{X})[1:n_2]^T \widehat{\mathbf{w}}^{n_2}_{MLE}\right] \neq \phi(\mathbf{X})[1:n_1]^T \mathbf{w}[1:n_1]$$

and that

$$Var\left[\phi(\mathbf{X})[1:n_2]^T \widehat{\mathbf{w}}^{n_2}_{MLE}\right] \leq Var\left[\phi(\mathbf{X})[1:n_1]^T \widehat{\mathbf{w}}^{n_1}_{MLE}\right]$$

(d) A consequence of excluding relevant basis functions in a linear model is that the predictions are unbiased, though they have larger variances. That is, for all integers $n_2 \in (0, n_1)$,

$$E\left[\phi(\mathbf{X})[1:n_2]^T \widehat{\mathbf{w}}^{n_2}_{MLE}\right] = \phi(\mathbf{X})[1:n_1]^T \mathbf{w}[1:n_1]$$

and that

$$Var\left[\phi(\mathbf{X})[1:n_2]^T \widehat{\mathbf{w}}^{n_2}_{MLE}\right] \geq Var\left[\phi(\mathbf{X})[1:n_1]^T \widehat{\mathbf{w}}^{n_1}_{MLE}\right]$$

(**10 Marks**)