ECE-UY 4563: Introduction to Machine Learning
Midterm 1 Solutions, Fall 2021

Prof. Sundeep Rangan

1. *Linear Regression.* Consider the model:

$$\widehat{y} = f(x) = \begin{cases} c_1 e^{-2x} + c_2 & \text{if } x < 1 \\ c_1 e^{-x} + c_3 & \text{if } x \geq 1. \end{cases} \tag{1}$$

for parameters $c = (c_1, c_2, c_3)$.

(a) Find **three** basis functions $\phi_1(x)$, $\phi_2(x)$, $\phi_3(x)$, such that every function in the model (1) can be written as:

$$\widehat{y} = \sum_{j=1}^{3} \beta_j \phi_j(x). \tag{2}$$

Write the β_j's in terms of the parameters c_j.

(b) Now find **two** basis functions, $\phi_1'(x)$ and $\phi_2'(x)$ for the model (1) with the additional constraint that $f(x)$ is continuous at $x = 1$.

Solution:

(a) We can take the basis functions,

$$\phi_1(x) = \begin{cases} e^{-2x} & \text{if } x < 1, \\ e^{-x} & \text{if } x \geq 1 \end{cases} \quad \phi_2(x) = \begin{cases} 1 & \text{if } x < 1, \\ 0 & \text{if } x \geq 1 \end{cases} \quad \phi_3(x) = \begin{cases} 0 & \text{if } x < 1, \\ 1 & \text{if } x \geq 1 \end{cases}$$

Then,

$$\widehat{y} = c_1\phi_1(x) + c_2\phi_2(x) + c_3\phi_3(x). \tag{3}$$

Hence, if we take $\beta_j = c_j$ the set of functions (1) is of the form (2).

(b) For continuity at $x = 1$ we need

$$\begin{aligned} \lim_{x\to 1^-} f(x) = \lim_{x\to 1^+} f(x) &\Rightarrow \lim_{x\to 1^-} (c_1 e^{-2x} + c_2) = \lim_{x\to 1^+} (c_1 e^{-x} + c_3) \\ &\Rightarrow c_1 e^{-2} + c_2 = c_1 e^{-1} + c_3 \\ &\Rightarrow c_3 = c_2 + c_1(e^{-2} - e^{-1}). \end{aligned}$$

So, we can write c_3 in terms of c_1 and c_2. This means that the expansion (3) can be

Model	Training MSE		Test MSE	
	Mean	SE	Mean	SE
1	10	1	10.1	1.1
2	4	0.4	4.1	0.5
3	3	0.3	4.5	0.6

Table 1: Problem 2: Training and test MSE results for the three models

rewritten as

$$\begin{aligned}\widehat{y} &= c_1\phi_1(x) + c_2\phi_2(x) + \left[c_2 + c_1(e^{-2} - e^{-1})\right]\phi_3(x) \\ &= c_1\left[\phi_1(x) + (e^{-2} - e^{-1})\phi_3(x)\right] + c_2\left[\phi_2(x) + \phi_3(x)\right].\end{aligned}$$

So, we can use two basis functions:

$$\phi_1'(x) = \phi_1(x) + (e^{-2} - e^{-1})\phi_3(x), \quad \phi_2'(x) = \phi_2(x) + \phi_3(x).$$

2. *Model Selection.* Consider three models for predicting a scalar y from $x = (x_1, x_2)$ where x_1 is a real valued variable and $x_2 \in \{1, 2, 3, 4\}$. Consider three models:

- Model 1: $\widehat{y} = b + wx_1$ where w is constant and does not depend on x_2.
- Model 2: $\widehat{y} = b + wx_1$ where w has one value when $x_2 = 1, 2$ and a different value when $x_2 = 3, 4$.
- Model 3: $\widehat{y} = b + wx_1$ where w has one of four values depending on x_2.

In all cases, the parameter b is constant and does not depend on x_1 or x_2.

Answer the following with short explanations (e.g., one sentence) for each of the following:

(a) Using a linear model for each case, what is the minimum number of parameters you need for each model?

(b) Which model would generally give the lowest *training* error?

(c) Which model would generally give the lowest *bias* error?

(d) Which model would generally give the lowest *variance* error?

(e) The results of K-fold validation are shown in Table 1. Which model would be selected based on the normal rule? Which model would be selected based on the one SE rule?

Solution:

(a) The number of parameters are as follows:

- Model 1: Two parameters: w and b.
- Model 2: Three parameters: b and two values of w.
- Model 3: Five parameters: b and four values of w.

(b) Model 3 since it is the most complex and can hence fit the data best.

(c) Model 3 since it is the most complex.

(d) Model 1 since it has the smallest number of parameters.

(e) For the normal rule, you would select Model 2 since it has the lowest test MSE. For the one SE rule, the target MSE is $4.1 + 0.5 = 4.6$. The simplest model with a mean MSE ≤ 4.6 is model 2. So, we would select model 2 for the one SE rule as well.

3. *Logistic Regression.* Consider a binary logistic classifier on a scalar x of the form

$$P(y = 1|x) = \frac{1}{1 + e^{-z}}, \quad z = b + wx.$$

(a) Find w and b such that

- $P(y = 1|x) = 0.8$ at $x = 3$ and
- $P(y = 0|x) = 0.8$ at $x = 1$.

(b) Using the values for w and b, what is $P(y = 1|x)$ at $x = 4$?

Solution:

(a) Since

$$P(y = 1|x) = \frac{1}{1 + e^{-z}},$$

we can solve for z:

$$z = -\ln\left[\frac{1}{P(y = 1|x)} - 1\right] = \ln\left[\frac{P(y = 1|x)}{1 - P(y = 1|x)}\right]. \tag{4}$$

Also, since $P(y = 0|x) = 1 - P(y = 1|x)$, we can write

$$z = \ln\left[\frac{1 - P(y = 0|x)}{1 - P(y = 0|x)}\right]. \tag{5}$$

Applying (4) at $x = 3$:

$$z = b + w(3) = \left[\frac{0.8}{1 - 0.8}\right] = \ln(4). \tag{6}$$

Applying (5) at $x = 1$:

$$z = b + w(1) = \left[\frac{0.2}{1 - 0.2}\right] = -\ln(4). \tag{7}$$

Subtracting (6) and (7),

$$2w = \ln(4) - (-\ln(4)) \Rightarrow w = \ln(4).$$

Then using (7),

$$b = -\ln(4) - w = -2\ln(4).$$

Hence, we have,

$$w = \ln(4) \approx 1.39, \quad b = -2\ln(4) \approx -2.77$$

(b) At $x = 5$,

$$z = wx + b = \ln(4)(4) - 2\ln(4) = 2\ln(4)$$

Hence,

$$\begin{aligned} P(y = 1|x) &= \frac{1}{1 + e^{-2\ln(4)}} \\ &= \frac{1}{1 + 4^{-2}} = \frac{16}{17} \approx 0.941 \end{aligned}$$