CSE 404: Introduction to Machine Learning (Fall 2020)
Homework #6
Due 11/9/2020 by 11.59pm

Note: (1) LFD refers to the textbook "Learning from Data". (2) Please upload a soft copy of your homework on D2L.

1. (50 points) **Exercise 3.6 (page 92) in LFD.** Cross-entropy error measure.

   (a) (25 points) More generally, if we are learning from ± 1 data to predict a noisy target $P(y|\mathbf{x})$ with candidate hypothesis h, show that the maximum likelihood method reduces to the task of finding h that minimizes

$$E_{in}(\mathbf{w}) = \sum_{n=1}^{N} [\![y_n = +1]\!] \ln \frac{1}{h(\mathbf{x}_n)} + [\![y_n = -1]\!] \ln \frac{1}{1 - h(\mathbf{x}_n)}$$

   Hint: Use the likelihood $p(y|x) = \begin{cases} h(x) & \text{for } y = +1 \\ 1 - h(x) & \text{for } y = -1 \end{cases}$ and derive the maximum likelihood formulation.

   (b) (25 points) For the case $h(\mathbf{x}) = \theta\left(\mathbf{w}^T \mathbf{x}\right)$, argue that minimizing the in-sample error in part (a) is equivalent to minimizing the one given below

$$E_{in}(\mathbf{w}) = \frac{1}{N} \sum_{n=1}^{N} \ln\left(1 + e^{-y_n \mathbf{w}^T \mathbf{x}_n}\right)$$

   Note from Book: For two probability distributions $\{p, 1-p\}$ and $\{q, 1-q\}$ with binary outcomes, the cross-entropy (from information theory) is

$$p \log \frac{1}{q} + (1-p) \log \frac{1}{1-q}.$$

   The in-sample error in part (a) corresponds to a cross-entropy error measure on the data point $(\mathbf{x}_n, y_n)$, with $p = [\![y_n = +1]\!]$ and $q = h(\mathbf{x}_n)$.

2. (50 points) **Exercise 3.7 (page 92) in LFD.** For logistic regression, show that

$$\begin{aligned} \nabla E_{in}(\mathbf{w}) &= -\frac{1}{N} \sum_{n=1}^{N} \frac{y_n \mathbf{x}_n}{1 + e^{y_n \mathbf{w}^T \mathbf{x}_n}} \\ &= \frac{1}{N} \sum_{n=1}^{N} -y_n \mathbf{x}_n \theta\left(-y_n \mathbf{w}^T \mathbf{x}_n\right) \end{aligned}$$

   Argue that a 'misclassified' example contributes more to the gradient than a correctly classified one.
   Hint: Remember the logistic regression objective function $E_{in}(\mathbf{w}) = \frac{1}{N} \sum_{n=1}^{N} \ln\left(1 + \exp\left(-y_n \mathbf{w}^T \mathbf{x}_n\right)\right)$ and take it's derivative with respect to $\mathbf{w}$.