Midterm 1 Practice for COMP 6321 Fall 2019

– Extensions

The questions in this practices midterm are suggestive only of the style and difficulty of questions that will be asked on on the real midterm. The length and the particular course content evaluated will be different.

Q1. [10 marks] This question is about logistic regression models
a) [2 marks] What kind of learning task is logistic regression used for?
b) [1 mark] Can the optimal parameter vector w for a logistic regression problem be solved for 'directly'?
c) [2 marks] Is the decision boundary of logistic regression linear or non-linear within the feature space Φ?
d) [3 marks] Assume you are given data set { (x_1, y_1), (x_2, y_2), … (x_N, y_N) }. Write the logistic regression loss function with respect to this data set. For full marks include the feature transformation $\Phi(.)$.

e) [2 marks] Assume you are given training set in matrix format $[1 \quad x_1 \quad x_2]$ where X and y are:

$$X = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -2 \\ 1 & 0 & 1 \\ 1 & 2 & 0 \end{bmatrix} \qquad y = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

Plot the data in two dimensions and draw the decision boundary that would result from applying logistic regression. Be sure to indicate which side corresponds to predicting y = 1.

Q2. [10 marks] Assume we have samples $\{x_1, x_2, \ldots, x_N\}$ from a univariate normal distribution $\mathcal{N}(\mu, \sigma)$. The likelihood $p(x \mid \mu, \sigma)$ having observed a single point x is therefore

$$\frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

a) [2 marks] The likelihood is a function of which variable(s)?

b) [2 marks] Write the likelihood $p(x_1, \ldots, x_N)$ having observed all x_i jointly.

c) [2 marks] Write the negative log likelihood of $p(x_1, \ldots, x_N \mid \mu, \sigma)$.

d) [2 marks] Write the gradient of the negative log likelihood of $p(x_1, \ldots, x_N \mid \mu, \sigma)$.

e) [2 marks] Use your answer from part (d) to derive a maximum likelihood estimate of the normal distribution parameters.

Q3. [8 marks] This question is about programming machine learning concepts with Numpy. You can assume that `import numpy as np` has already been run.

a) [2 marks] You are given the following incomplete function:

```
def linear_model_predict(X, w):
    """
    Returns predictions from linear model y(X, w) at each point X[i, :] using parameters w.
    Given X with shape (N,D+1) and w must have shape (D+1,), and return result will have shape (N,).
    """
```

Complete the function in the space below. (No need to copy the above function signature.) For full marks, your answer should be fully vectorized.

b) [2 marks] You are given the following incomplete function:

```
def sigmoid(z):
    """
    Return the element-wise logistic sigmoid of array z.
    """
```

Complete the function in the space below (No need to copy the above function signature.) For full marks, your answer should be fully vectorized.

c) [4 marks] You are given the following incomplete function:

```
def linear_regression_by_gradient_descent(X, y, w_init, learn_rate=0.05, num_steps=500):
    """
    Fits a linear model by gradient descent.

    If the feature matrix X has shape (N, D), the targets y should have shape (N,)
    and the initial parameters w_init should have shape (D,).

    Returns a new parameter vector w that minimizes the squared error to the targets.
    """
```

The gradient of a linear model can be expressed mathematically as

$$\nabla \ell_{\text{LS}} = (X^T X)w - X^T y$$

Complete the function in the space below.

d) [4 marks] You are given the following incomplete function:

```
def linear_regression_by_direct_solve(X, y):
    """
    Fits a linear model by directly solving for the optimal parameter w.
    """
```

The gradient of a linear model can be expressed mathematically as

$$\nabla \ell_{\text{LS}} = (X^T X)w - X^T y$$

Complete the function in the space below.

e) [4 marks] You are given the following incomplete function:

```
def logistic_model_predict(X, w):
    """
    Returns predictions from logistic model y(x, w) at each point X[i, :] using
parameters w.
   Given X with shape (N, D+1), w must have shape (D+1,) and the result will have
shape (N,).
    """
```

Complete the function in the space below.

f) [4 marks] You are given the following incomplete function:

```
def logistic_regression_grad(X, y, w):
    """
   Returns the gradient for basic logistic regression.
    """
```

The basic logistic regression training objective is:

$$\ell_{\text{LR}}(\mathbf{w}) = \sum_{i=1}^{N}[\, y_i \ln \sigma(w^T x_i) + (1 - y_i) \ln\big(1 - \sigma(w^T x_i)\big)\,]$$

Complete the function in the space below.

g) [4 marks] You are given the following incomplete function:

```
def logistic_regression(X, y, w_init, learn_rate=0.05, num_steps=500):
    """
     Fits a logistic model by gradient descent.


    If the feature matrix X has shape (N,D), the targets y should have shape (N,)
and the initial parameters w_init should have shape (D,).


   Returns a new parameter vector w that minimizes the negative log likelihood of
the targets
    """
```

The basic gradient for the above training objective is:

$$\nabla \ell_{\text{LR}}(\mathbf{w}) = \sum_{i=1}^{N}(\sigma(w^T x_i) - y_i)\, x_i$$

Complete the function in the space below.

Q4. [6 marks] This question is about the assigned reading: the 2001 paper by Leo Breiman.

a) [2 marks] After having worked as a consultant, what were Breiman's "perceptions" about how to work with data?

b) [1 marks] How was predictive accuracy measured in Breiman's "Ozone project"?

c) [2 marks] Describe the modeling approach that Breiman's team used in the "Chlorine project"

d) [2 marks] Describe an example where theory in algorithmic modelling led to an important advance.

e) [2 marks] What learning algorithm does Breiman rate as “A+ for prediction” but “F for interpretability”, and what are his reasons?

f) [2 marks] State ‘Occam’s dilemma’ as Breiman describes it.

g) [2 marks] Describe the main symptom of ‘model instability’.

h) [2 marks] What is the ‘straight jacket’ that Breiman claims statisticians are imposing on themselves? Why does it matters?