Homework 3
MSCS 550 | CMPT 404

Pablo Rivas

Assigned: Sep/27/16; Due: Oct/11/16; Points: 60

1 Instructions

This assignment could be written in LaTeX, just as the last homework assignment. Write in understandable, easy to follow English. Make sure you provide good illustrations and figures. Remember to include your Python programs in your assignment.

Your assignment should be submitted in two ways: through GitHub, and in hardcopy (in class). Use the **same** repository you have been using and submit your work in a folder named "`lastname-xx`", where lastname is your last name xx is the number of the assignment.

2 Problem Set

The following is a list of problems you will work on. When providing your solutions (hopefully using LaTeX), do not simply give the final answer, show how you arrived to the solution, justify your assumptions, and explain your results clearly.

1. Compare two algorithms on a classification task: the Pocket algorithm (designed for classification), and linear regression (not designed for classification). For linear regression, after learning the weights $\mathbf{w}$; we use $h(\mathbf{x}) = \text{sign}(\mathbf{w}^T\mathbf{x})$ to classify $\mathbf{x}$. For the dataset, start by using the following Python code:

```python
import numpy as np
import matplotlib.pyplot as plt
from sklearn.datasets.samples_generator import make_blobs
# generates random centers
ctrs = 3 * np.random.normal(0, 1, (2, 2))
# generates random data following normal distributions
X, y = make_blobs(n_samples=100, centers=ctrs, n_features=2,
         cluster_std=1.0, shuffle=False, random_state=0)
y[y==0] = -1     #makes sure we have +1/-1 labels
# plots data
c0 = plt.scatter(X[y==-1,0], X[y==-1,1], s=20, color='r', marker='x')
c1 = plt.scatter(X[y==1,0], X[y==1,1], s=20, color='b', marker='o')
# displays legend
plt.legend((c0, c1), ('Class -1', 'Class +1'), loc='upper right',
         scatterpoints=1, fontsize=11)
# displays axis legends and title
plt.xlabel(r'$x_1$')
plt.ylabel(r'$x_2$')
plt.title(r'Two simple clusters of random data')
# saves the figure into a .pdf file (desired!)
plt.savefig('hw3.plot.pdf', bbox_inches='tight')
plt.show()
```

Create another dataset using the same methods as above, which we will use to estimate E_{out}.

Try the following three approaches using multiple randomized experiments and explain which works best in terms both E_{out} and the amount of computation required.

(a) The Pocket algorithm, starting from $\mathbf{w} = 0$.

(b) Linear regression (applied as a classification method).

(c) The Pocket algorithm, starting from the solution given by linear regression.

Also, try adding some *significant* outliers to the $y = +1$ class (arbitrarily chosen) of the training dataset and explain how that affects your results.

2. (graduate students) Consider the logistic regression model and its likelihood function:

$$\sigma(\alpha) = \frac{1}{1 + e^{-\alpha}} \tag{1}$$

$$P(y = 1|\mathbf{x}) = \sigma(\mathbf{w}^T\mathbf{x}) \tag{2}$$

$$P(y = -1|\mathbf{x}) = 1 - \sigma(\mathbf{w}^T\mathbf{x}) \tag{3}$$

$$P(y|\mathbf{x}) = P(y = 1|\mathbf{x})^{\frac{y+1}{2}} P(y = -1|\mathbf{x})^{\frac{1-y}{2}} \tag{4}$$

$$\ell(\mathbf{w}) = \log \prod_{n=1}^{N} P(y_n|\mathbf{x}_n) = \sum_{n=1}^{N} \log P(y_n|\mathbf{x}_n) \tag{5}$$

(a) Show that

$$\frac{d\sigma}{d\alpha} = \sigma(\alpha)(1 - \sigma(\alpha)). \tag{6}$$

(b) Derive the gradient of the log-likelihood, $\nabla_{\mathbf{w}}\ell(\mathbf{w})$.

(c) Write down the update step for gradient ascent of $\ell(\mathbf{w})$ using the gradient you just derived.