CSE446 Machine Learning, Winter 2016: Homework 2

Due: Monday, February 8^{th}, beginning of class

Start Early! Also, typed solutions (specifically those in LaTeX) are preferred to hand-written solutions. Any illegible solutions will be counted wrong at the sole discretion of the grader. Please feel free to use the homework document as a template, putting your solutions inline. We will only accept answers in **.pdf** format.

1 Naive Bayes [50 points]

In this problem we'll use Naive Bayes to predict whether a person has a cold given the symptoms $Headache$, $Cough$, $SoreThroat$. Imagine we have observed the following data:

Disease	Headache	Cough	Sore Throat
$Cold$	T	T	T
$Cold$	T	T	T
$Cold$	T	F	F
$\neg Cold$	F	F	F
$\neg Cold$	F	F	F
$\neg Cold$	T	T	T

1. (10 points) Remember that when using Naive Bayes, we make the assumption that all features are conditionally indepedent given the target value. Factorize the joint probability $P(Cold, Headache, Cough, SoreThroat)$ using this assumption.

2. (10 points) Using no Laplacian smoothing, what is $P(Cold|\neg Headache, Cough, SoreThroat)$?

3. (15 points) Now estimate the conditional likelihoods using Laplacian smoothing with $\alpha = 1$. What is $P(Cold|Headache, \neg Cough, \neg SoreThroat)$?

4. (15 points) In our dataset above, $Cough$ and $SoreThroat$ are completely correlated, which doesn't fit with our conditional independence assumption. Ignore the $SoreThroat$ feature, and compute $P(Cold|Headache, \neg Cough)$ (use Laplacian smoothing with $\alpha = 1$ again). Is our prediction the same as when we used the $SoreThroat$ feature?

2 Perceptrons [50 points]

Recall that a perceptron learns a linear classifier with weight vector w. It predicts

$$y^{(i)} = \text{sign}(w^T x^{(i)})$$

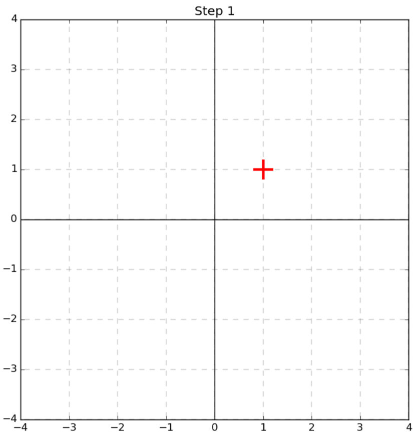
(assuming here that $y^{(i)} \in \{+1, -1\}$. Also, note that we are *not* using a bias weight w_0, for simplicity). When the perceptron makes a mistake, it updates the weights using the formula

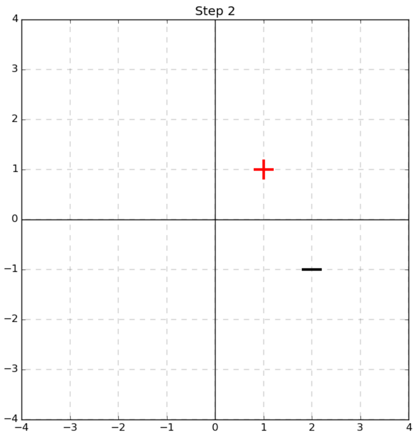
$$w = w + y^{(i)} x^{(i)}$$

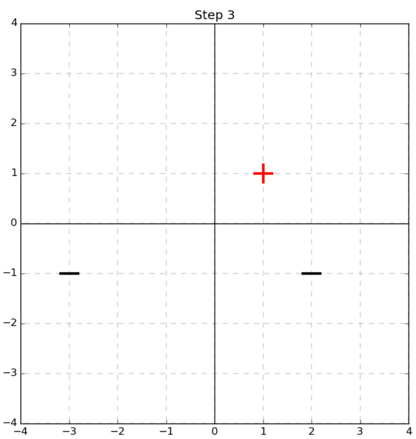
Imagine that we have $x^{(i)} \in \mathbb{R}^2$, and we encounter the following data points

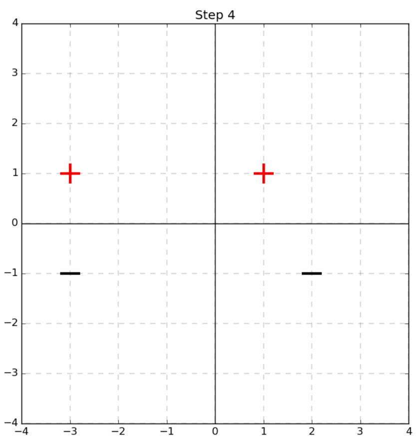
x_1	x_2	y
1	1	1
2	-1	-1
-3	-1	-1
-3	1	1

1. (25 points) Starting with $w = [0 \quad 0]^T$, use the perceptron algorithm to learn on the data points in the order from top to bottom. Show the perceptron's linear decision boundary after observing each data point in the graphs below. Be sure to show which side is classifed as positive.

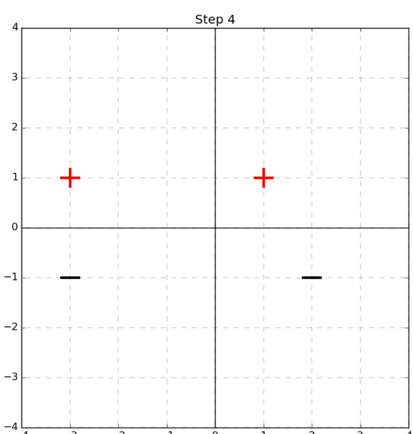








2. (10 points) Does our learned perceptron maximize the margin between the training data and the decision boundary? If not, draw the maximum-margin decision boundary on the graph below.



3. (15 points) Assume that we continue to collect data and train the perceptron. If all data we see (including the points we just trained on) are linearly separable separable with margin $\gamma = 0.5$ and have maximum norm $||x^{(i)}|| \leq 5$, what is the maximum number of mistakes we can make on future data?

3 Logistic Regression [100 points]

The programming portion of this assignment can be found at `https://github.com/pjreddie/Logistic-SGD`.