CSE 151A - Discussion 05

Quick Review

Logistic Regression
Predict a probability $H_{\vec{w}}(\vec{x}) = \sigma(\vec{w} \cdot Aug(\vec{x}))$, with logistic function $\sigma(t) = \frac{1}{1+e^{-t}}$
Goal : find the value of $\vec{w}$ to maximize the log-likelihood $f(\vec{w}) = log\mathcal{L}(\vec{w})$ using gradient ascent

Maximum Likelihood
In general, the likelihood is : $\mathcal{L}(\vec{w}) = \prod_{i=1}^{n} \frac{1}{1 + e^{-y_i \vec{w} \cdot Aug(\vec{x}^{(i)})}}$

The log likelihood is : $f(\vec{w}) = log\mathcal{L}(\vec{w}) = -\sum_{i=1}^{n} log\left[1 + e^{-y_i \vec{w} \cdot Aug(\vec{x}^{(i)})}\right]$

*NOTE : Here we have assumed $y_i = 1$ for a positive class label, and $y_i = -1$ for a negative class label. You may encounter equations that look different to the ones used here, and this is likely due the use of $y_i = 0$ rather than $y_i = -1$ for a negative class label.

Gradient Ascent
Setting : $f(\vec{w})$ is differentiable but we cannot explicitly solve for $\vec{w}$ like before
Strategy : pick a starting guess $\vec{w}^{(0)}$ and iterate $\vec{w}^{(i)} = \vec{w}^{(i-1)} + \alpha \cdot \nabla f(\vec{w}^{(i-1)})$ until convergence, where $\nabla f(\vec{w}^{(i-1)}) = \sum_{k=1}^{n} y_k \vec{x}^{(k)} H_{\vec{w}^{(i-1)}}(-y_k \vec{x}^{(k)})$

Making Classifications
Predict class 1 if $H_{\vec{w}}(\vec{x}) > \tau$ $\quad$ τ can be thought of as a threshold probability (y-value of logistic function)
Predict class 1 if $\vec{w} \cdot Aug(\vec{x}) > t$ $\quad$ t can be thought of as an x-value threshold on the logistic function

Problem 1.

Consider the equation $f(w) = -(x^2 - 5x + 4)$.

a) Show that this function is strictly concave. What does this tell us about the number of maxima?

b) Use gradient ascent to solve for the value of x that maximizes $f(x)$.
For this problem, start with $x^{(0)} = 0$, and compute all intermediate $x^{(i)}$ values up to $x^{(4)}$.
Do this three times with the following values of $\alpha : 0.3, 0.8, 1.2$. For each, note any interesting findings and determine if the algorithm will eventually converge.

Problem 2.

After running gradient descent, suppose that we have solved for $\vec{w} = (0.5, 2, -1)^T$ that minimizes some convex function f.
We have the following validation set, consisting of four data points and their corresponding labels:

i	$x_1^{(i)}$	$x_2^{(i)}$	y_i
1	2	4	-1
2	3	2	1
3	0	-1	-1
4	-1	2	-1

* Note : Don't forget to augment each $\vec{x}^{(i)}$

a) We will use the following rule : Predict class 1 if $H_{\vec{w}}(\vec{x}) > \tau$, else predict class -1.
What is the classification accuracy over the above validation set when $\tau = 0.5$?

b) Is there a value of τ that would result in 100% validation accuracy? If so, compute such a value.