Stat 344ne - Potential Quiz 1 Topics/Things to Know

Here are some things that I may ask you to know how to do on a quiz on Friday. I'll answer questions about any of these things (or anything else) on Monday and Wednesday.

- Draw a diagram of a neural network corresponding to logistic regression
- Explain what it means if I write $Y^{(i)} \sim \text{Bernoulli}(a^{(i)})$ in the context of specifying a logistic regression model:
  - $Y^{(i)}$ is a random variable – as far as our model is concerned, its exact value is unknown
  - The possible values of $Y^{(i)}$ are either 0 or 1
  - The model's estimated probability that $Y^{(i)} = 1$ is $a^{(i)}$
- Show (in 2 equations, one for $z^{(i)}$ and a second for $a^{(i)}$) how a logistic regression model can be viewed as calculating $P(Y^{(i)} = 1)$ by first calculating $z^{(i)}$ and second calculating $a^{(i)}$. These equations are for a single observation, number i.
- Write down an equation for the sigmoid function. Explain why it is between 0 and 1 (for example you could argue that the numerator is smaller than the denominator so it must be < 1 and the numerator and denominator are both positive so it must be > 0). Be able to draw a picture of the sigmoid function.
- Show that the decision boundary for a basic logistic regression model with no non-linear functions of x is linear
- Explain how you could get a non-linear decision boundary (such as an elliptical decision boundary) by modifying the basic logistic regression model above
- Write down the following equation for how you could calculate the vector z for all observations $1, \ldots, m$:

$$\begin{bmatrix} z^{(1)} \\ z^{(2)} \\ \vdots \\ z^{(m)} \end{bmatrix} = \begin{bmatrix} b \\ b \\ \vdots \\ b \end{bmatrix} + \begin{bmatrix} \left(x^{(1)}\right)^T \\ \left(x^{(2)}\right)^T \\ \vdots \\ \left(x^{(m)}\right)^T \end{bmatrix} w$$

- Also be able to write down a version of the above equation where everything is transposed
- In the above equations, what does $x^{(1)}$ represent? What does w represent? What do we call w and b
  - $x^{(1)}$ is a column vector of features for the first observation in our data set
  - w is a column vector of *weights* corresponding to each feature
  - b is a bias; it plays the role of an intercept term
- What is the main idea of maximum likelihood estimation?
  - We choose the model parameters so that the probability of the observed data is as large as possible. In logistic regression, the model parameters to estimate are w and b.
- For logistic regression, what does the likelihood function look like, specified in terms of $a^{(i)}$ and $y^{(i)}$? Do we want to maximize this or minimize it?
  - $\prod_{i=1}^{m} \left\{a^{(i)}\right\}^{y^{(i)}} \left\{1 - a^{(i)}\right\}^{1-y^{(i)}}$
  - We want to maximize this
- If I give you three specific values of $a^{(i)}$ and $y^{(i)}$ for three observations $i = 1, 2, 3$, be able to actually use the expression above to calculate the value of the likelihood function
- For logistic regression, what does the log-likelihood function look like, specified in terms of $a^{(i)}$ and $y^{(i)}$? Do we want to maximize this or minimize it? (**I realize I didn't write this down in class on Friday, we'll take 2 minutes to do that on Monday**)
- For logistic regression, what does the negative log-likelihood function look like, specified in terms of $a^{(i)}$ and $y^{(i)}$? Do we want to maximize this or minimize it? (**I realize I didn't write this down in class on Friday, we'll take another 2 minutes to do that on Monday**)