4.3 class notes
Kastan Day
February 15, 2019

1 Logistic Regression

Logistic functions to make estimates of binary and multi-class problems.

Binary case \((y \in 0, 1)\):

- \(h_{\vec{b}} = p(y = 1|\vec{x}, \vec{b})\)
- \(h_{\vec{b}} = g(\vec{b}^T \vec{x})\) ... The inner part is our linear model

\[ g(z) = \frac{1}{1+e^{-z}} \] ... This is the logistic sigmoid

This is a logistic sigmoid graph:

Goal: Find \(\vec{b}\) by maximizing the likelihood of our data

Aside to likelihoods:
Flipping a weighted coin (Bernoulli Random Variable).
- \(n\) coin flips
- \(p\) probability of heads
- \(H = 1, T = 0\).
n = 10, $\bar{y} = 0, 0, 1, 0, 1, 0, 1, 0, 0$
Likelihood, $L(p) = (1-p)p(1-p)p ... (1-p)$
L(p) = $(1 - p)^{\bar{y}}p^{\bar{y}}$ ... for our specific example
In general $L(P) = \prod_{i=1}^{n} (p)^{y_i}(1 - p)^{1-y_i}$

Logistic Regression:
The likelihood of B with respect to our data, for logistic regression. We want to maximize this likelihood.
In general $L(P) = \prod_{i=1}^{n} h_{\vec{b}}(\vec{x}_i)^{y_i}(1 - h_{\vec{b}}(\vec{x}_i))^{1-y_i}$
Note: First term is Prob that $Y = 1$, and the second term is the prob that $Y = 0$, for a specific example.

We're going to maximize the log likelihood, which will have the same maximization as the function itself, that's why it's okay and common to max log likelihoods instead of the function itself.

$l(P) = log(L(P)) = \sum_{i=1}^{n} y_i log(p) + \sum_{i=1}^{n} log(1 - p)$

Simplify, take the derivative of this with respect to p, and set it equal to 0. Then solve for P.

$\frac{n\bar{y}}{p} - \frac{n-n\bar{y}}{1-p}$
... and assume n=1 (or is a constant) I think?
$\bar{y} = p... p$ is actually $\hat{p}$ in the end

In the dice example from above:
$\bar{y} = \frac{0+6+4+1}{10}$
$\hat{p} = \bar{y} = \frac{2}{5} = 0.4$