3032247297

IEOR 142 HW #2

(1) a) It makes more sense to apply a log regression model to a seperable dataset because if we use seperable data we risk a logistical regression model that doesn't converge, which would prevent us from being able to improve our model.

b) $f(t) = \log(1+e^{-t})$

$g(t) = \log(1+e^{t}) - t$

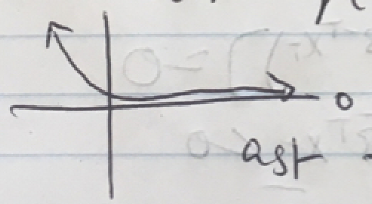
$t = \log(e^{t})$

$\log\left(1+\frac{1}{e^{t}}\right) = \log(1+e^{t}) - \log(e^{t})$

$= \log\left(\frac{e^{t}}{e^{t}} + \frac{1}{e^{t}}\right) = \log\left(\frac{1+e^{t}}{e^{t}}\right)$

$= \log\left(\frac{1+e^{t}}{e^{t}}\right) = \log\left(\frac{1+e^{t}}{e^{t}}\right)$ ✓

c) $f(t) = \log(1+e^{-t})$



as $t \to \infty$ $f(t) = 0$

... that we did

d) $\min_{\beta} \left\{ \sum^{n} \left[\log\left(1+e^{(\beta^T x_i)}\right) - y_i \beta^T x_i \right] \right\}$

$\rightarrow \lim_{t \to \infty} \log(\text{loss}(t\bar{\beta})) = 0$

i) $y_i = 1$

$\lim_{t \to \infty} \sum \left[\log\left(1+e^{\infty \bar{\beta}^T x_i}\right) - \beta^T x_i \right] = 0$

we know

$\log(1+e^{-t}) = \log(1+e^{t}) - t \Rightarrow$

$\log\left(1+e^{-(t\bar{\beta} x_i)}\right) = \log\left(1+e^{t\bar{\beta} x_i}\right) - t\bar{\beta}^T x_i$

Above is true when $\beta^T x_i > 0$

$a_1 x_{i1} \dots x_n x_{in} - \infty > 0$ for $y_i = 1$

so $\bar{\beta}$ exists such that $\bar{\beta}^T x_i > 0$

ii) $y_i = 0$

$\lim_{t \to \infty} \sum \left[\log\left(1+e^{(t\bar{\beta} x_i)}\right) \right]$

$\lim_{t \to \infty} \sum \left[\log\left(1+e^{\infty \beta^{-T} x_i}\right) \right] = 0$

Above is true if $\bar{\beta}^T x_i < 0$

so there exists a $\bar{\beta}$ such that

$\bar{\beta}_1 x_{i1} + \dots \bar{\beta}_n x_{in} < 0$

1e) we saw from the work that we did that β can always increase which means that our loss function can approach 0. This means that our error can always decrease, which means that the accuracy and quality of our model can increase. Since there are infinite values of β that can make our model better our $f(x)$ will never converge, therefore it makes more sense to work w/ a non seperable dataset bc seperable datasets converge. Ultimately the bad behavior of the optimization problem does align with this intuition.