Homework 3
Problem 2
Solution
a) Logistic Function
$\sigma(\alpha) = \frac{1}{1+e^{(-\alpha)}}$
$\sigma(\alpha) = \frac{e^{\alpha}}{1+e^{\alpha}}$
$= (e^{\alpha})(1+e^{\alpha})^{-1}$
Chain Rule
$\frac{d\sigma}{d\alpha} = (e^{\alpha})(1+e^{\alpha})^{-1} + (e^{\alpha})(1+e^{\alpha})^{-2}(e^{\alpha})$
$\frac{d\sigma}{d\alpha} = \frac{(e^{\alpha})(1+e^{\alpha})}{(1+e^{\alpha})^2} - \frac{e^{\alpha 2}}{(1+e^{\alpha})^2}$
$= \frac{(e^{\alpha})}{(1+e^{\alpha})^2}$
$= \frac{(e^{\alpha})}{(1+e^{\alpha})} * \frac{1}{(1+e^{\alpha})}$
$= \sigma(\alpha) * \sigma(-\alpha)$
$= \sigma(\alpha)(1-\sigma(\alpha))$
b)
Log function:
$l(w) = \log \prod_{n=1}^{N} P(y_n|x_n)$
Re-writing the equation
$l(w) = \prod_{n=1}^{N} P(y_n|x_n)$
Therefore
$P(y_n|x_n) = P(y_n = +1|x_n) * P(y_n = -1|x_n)$
$l(w) = \Sigma_{n=1}^{N} \log(P(y_n = +1|x_n) * P(y_n = +1|x_n))$
$l(w) = \Sigma_{n=1}^{N} \log(P(y_n = +1|x_n) * (1 - P(y_n = +1|x_n)))$
$l(w) = \Sigma_{n=1}^{N}([y_n = +1] \log P(y_n = +1|x_n) + [1-(y_n = +1)] \log(1-P(y_n = +1|x_n))$
$l(w) = \Sigma_{n=1}^{N}[[y_n = +1] \log(\frac{1}{1+e^{-w^T x_n y_n}}) + [1-(y_n = +1)] \log(\frac{e^{-w^T x_n y_n}}{1+e^{-w^T x_n y_n}})]$
$l(w) = \Sigma_{n=1}^{N}[-[y_n = +1] log(1+e^{-w^T x_n y_n}) + [1-(y_n = +1)]((-w^T x_n y_n) - \log(1+e^{-w^T x_n y_n}))]$
$l(w) = \Sigma_{n=1}^{N}[-[1-(y_n = +1)](-w^T x_n y_n) - \log(1+e^{-w^T x_n y_n})]$
Derivative W.r.t w:
$\frac{dl(w)}{dw} = \Sigma_{n=1}^{N}[-[1-(y_n = +1)]\frac{d}{dw}(-w^T x_n y_n) - \frac{d}{dw}(1+e^{-w^T x_n y_n})]$
$\frac{dl(w)}{dw} = (x_n y_n)\Sigma_{n=1}^{N}[[1-(y_n = +1)] + P(y_n = -1|x_n)]$
$\frac{dl(w)}{dw} = (-x_n y_n)\Sigma_{n=1}^{N}[[(y_n = +1)] - P(y_n = -1|x_n)]$
After Averaging:
$\Delta_w l(w) = \frac{dl(w)}{dw} = \frac{1}{N}(-x_n y_n)\Sigma_{n=1}^{N}[[(y_n = +1)] - P(y_n = -1|x_n)]$

$\Delta_w l(w) = \frac{1}{N}(-x_n y_n)\Sigma_{n=1}^N \sigma(-y_n w^T x_n)$
$\Delta_w l(w) = \frac{1}{N}(-y_n x_n)\Sigma_{n=1}^N \sigma(-y_n w^T x_n)$
(c)
Taylor expression:
$\Delta_w l(w) = l(w)(w(0) + \eta \upsilon) - l(w(0))$
$= \eta(w)(w(0))^T \upsilon + O(\eta)^2$
$\geq -\eta||(w)(w(0))^T||$
Deriving the update step
$e_n(w) = ln(1 + e^{-y_n w^T I_n}$
${}_n(w) = \frac{-y_n x_n}{1+e^{y_n w^T x_n}}$
Updated weight:
$w \leftarrow w - \eta_n(w)$
Expected weight:
$-\eta \frac{1}{N} \Sigma_{n=1}^N {}_n(w)$