Practice Midterm.

Problem 1 (a) $y = \sigma(z) = \frac{1}{1+e^{-z}}$

$\frac{\partial}{\partial z}\frac{1}{1+e^{-z}} \Rightarrow g = \frac{u}{v}\ ;\ g' = \frac{u'v - v'u}{v^2}$

$u = 1,\ v = 1+e^{-z},\ v' = -e^{-z},\ u' = 0.$

$\Rightarrow \frac{u'v - v'u}{v^2} = \frac{0-(-e^{-z})\times 1}{(1+e^{-z})^2} = \frac{e^{-z}}{(1+e^{-z})^2}$

$y(1-y) = y - y^2 = \frac{1}{1+e^{-z}} - \frac{1}{(1+e^{-z})^2} = \frac{1+e^{-z}-1}{(1+e^{-z})^2} = \frac{e^{-z}}{(1+e^{-z})^2}$

(b) $\frac{\partial}{\partial w_j} w^T x + b = \frac{\partial}{\partial w_j} w_1x_1 + \cdots + w_jx_j + \cdots + w_dx_d + b$

$= \frac{\partial}{\partial w_j}\overset{0}{\cancel{w_1x_1}} + \cdots + \frac{\partial}{\partial w_j}\overset{x_j}{\cancel{w_jx_j}} + \cdots + \frac{\partial}{\partial w_j}\overset{0}{\cancel{w_dx_d}} + \frac{\partial}{\partial w_j}\overset{0}{\cancel{b}}$

$= x_j$

(c). $\frac{\partial}{\partial b} w^T x + b = \frac{\partial}{\partial b}\overset{0}{\cancel{w_1x_1}} + \cdots + \frac{\partial}{\partial b}\overset{0}{\cancel{w_dx_d}} + \frac{\partial}{\partial b}\overset{1}{\cancel{b}} = 1$

Problem 2 (a) A - KL divergence

(b) Among all the choices, only KL divergence is not symmetrical.

Problem 3. Gradient descent (all its variation included) can only find local minimum, which means it's not always ~~not~~ necessarily true that gd will find the global minimum (and thus finding the configuration of the parameters that gives us the global minimum

Problem 4. (c) the "k" represents the # of nearest neighbors in KNN.

Problem 5. $w_j \leftarrow w_j - \partial_{w_j} \nabla J_{reg} \Leftrightarrow w_j \leftarrow w_j - \partial w_j \cdot \frac{\partial J_{reg}}{\partial w_j}$

$b \leftarrow b - \partial_b \nabla J_{reg} \Leftrightarrow b \leftarrow b - \partial b \cdot \frac{\partial J_{reg}}{\partial b}$

Problem 6. $x^{(a)} = [5, 9, -3]$, $x^{(b)} = [1, 2, -6]$

Euclidean$(x^{(a)}, x^{(b)}) =$

$\sqrt{(x_1^{(a)} - x_1^{(b)})^2 + (x_2^{(a)} - (x_2^{(b)}))^2 + (x_3^{(a)} - x_3^{(b)})^2}$

$= \sqrt{((5-1)^2 + (9-2)^2 + (-3+6)^2)} = \sqrt{74} \approx 8.60$

Problem 7. w_0 = bias incorporated within the weight matrix. This value/~~is~~ variable is essentially the bias.

w_1 = the weight associated with the 1st feature within my sample.

Remark: w_j = the weight associated with the j^{th} feature

Problem 8 (a) this sample is not a support vector and has no say in decididing the margin

(b) this sample is a support vector and has some say in deciding the margin

(c) No.

(d) a negative Laugraigian multiplier is saying that this sample negatively contributes to the maximizing of the margin (i.e. take ~~in~~ it as a support vector would make the margin smaller, not bigger).

Problem 9. $x = \{1, 2, 3, 4\}$ $W_{iter=0} = \{0.1, 0.1\}$.

$t = \{10, 20, 30, 40\}$

$y_1 = 0.1 + 0.1 \cdot 1 = 0.2$ $y_4 = 0.1 + 0.1 \cdot 4 = 0.5$

$y = w_0 + w_1 \cdot x$ $y_2 = 0.1 + 0.1 \cdot 2 = 0.3$

$y_3 = 0.1 + 0.1 \cdot 3 = 0.4$

$y = \{0.2, 0.3, 0.4, 0.5\}$

$J(w_0, w_1) = \frac{1}{N} \cdot \sum_1^N \mathcal{L}_i = \frac{1}{N} \cdot \sum_1^N \frac{1}{2}(y_i - t_i)^2$

$\frac{\partial \mathcal{L}}{\partial w_0} = \frac{\partial \mathcal{L}}{\partial y} \cdot \frac{\partial y}{\partial w_0}$; $\frac{\partial \mathcal{L}}{\partial w_1} = \frac{\partial \mathcal{L}}{\partial y} \cdot \frac{\partial y}{\partial w_1}$

$\frac{\partial \mathcal{L}}{\partial y} = \frac{\partial}{\partial y} \frac{1}{2}(y_i - t_i)^2 = y_i - t_i$

$\frac{\partial y}{\partial w_0} = \frac{\partial}{\partial w_0} w_0 + w_1 x = 1$; $\frac{\partial y}{\partial w_1} = \frac{\partial}{\partial w_1} w_0 + w_1 x = x$

$\Rightarrow$ $\nabla J(w_0) = \frac{1}{N} \cdot \sum_1^N \nabla \mathcal{L}_i(w_0)$

$= \frac{1}{N} \cdot \sum_1 (y_i - t_i) \cdot 1$

$= \frac{1}{4}((0.2 - 10) + (0.3 - 20) + (0.4 - 30) + (0.5 - 40))$

$= \frac{1}{4}(-98.6) = -24.65$

$\nabla J(w_1) = \frac{1}{N} \sum_1 \nabla \mathcal{L}_i(w_1)$

$= \frac{1}{N} \sum_1 (y_i - t_i) \cdot x$

$= \frac{1}{4}((0.2 - 10) \cdot 1 + (0.3 - 20) \cdot 2 + (0.4 - 30) \cdot 3 + (0.5 - 40) \cdot 4))$

$= \frac{1}{4}(-296.0) = -74.0$

$w_0 \leftarrow w_0 - \alpha \cdot \nabla J(w_0) = 0.1 - 0.1 \cdot (-24.65) = 2.565$

$w_1 \leftarrow w_1 - \alpha \cdot \nabla J(w_1) = 0.1 - 0.1 \cdot (-74.0) = 7.5$

$W_{iter=1} = \{2.565, 7.5\}$