Problem Set 8

8.5

a) Since $f(x_i) = w^T x_i$, $Z_i = y_i f(x_i)$ if the classifing is correct

$y_i = f(x_i)$ and they have the same sign, which means the correct classify $\Rightarrow$ $y_i f(x_i) \geq 0$

Futher more : classify correctly : $Z_i \geq 0$

classify wrongly : $Z_i < 0$

Thus the number of misclassified training points = the number of $Z_i < 0$, to minimize it

$$\Rightarrow \quad R_{emp}(w) = \sum_{i=1}^{n} L_{01}(Z_i) , \quad \text{where } L_{01}(Z_i) = \begin{cases} 0 & Z_i \geq 0 \\ 1 & Z_i < 0 \end{cases}$$

b) To the perceptron, it only look at misclassified points.

And the misclassified points mean $Z_i < 0$

$E(w) = \sum_{i \in \mathcal{U}} -Z_i$ and to $\mathcal{U} = \{\text{misclassified points}\}$

$= \{ i \mid Z_i = y_i w^T x_i < 0 \}$

Since don't need care correct points.

just $Z_i < 0$, use $-Z_i$

$$\text{Thus,} \quad E(w) = \sum_{i \in \mathcal{U}} -Z_i = \sum_{i=1}^{n} I(Z_i) , \quad \text{where } I(Z_i) = \begin{cases} 0 & Z_i \geq 0 \\ -Z_i & Z_i < 0 \end{cases}$$

$I(Z_i) \Rightarrow \max(0, -Z_i)$

So $E(w) = \sum_{i=1}^{n} \max(0, -Z_i)$, where $L_p(Z_i) = \max(0, -Z_i)$

c) To the Least-Squares

$$\hat{w} = \underset{w}{\operatorname{argmin}} \sum_{i=1}^{n} \| w^T x_i - y_i \|^2 = \underset{w}{\operatorname{argmin}} \sum_{i=1}^{n} \| y_i w^T x_i - y_i^2 \|^2$$

and $y_i^2 = 1$, $y_i w^T x_i = Z_i$

Thus, $R_{emp} = \sum_{i=1}^{n} (z_i - 1)^2$, $L_{LSC}(z_i) = (z_i - 1)^2$

d) To the logistic regression, $\sigma(w^T x_i) = \pi_i = p(y|x)$

Since $y^* = \begin{cases} 1 & p(y|x) > \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$ from $\sigma(w^T x_i) = \frac{1}{1+e^{-w^T x_i}}$

Then, map the $y^* = \{0, 1\}$ to $y = \{-1, 1\}$

$$E(w) = -\sum_{i=1}^{n} y_i^* \log \pi_i + (1 - y_i^*) \log (1 - \pi_i)$$

$$= -\sum_{i=1}^{n} \begin{cases} \log \pi_i & y_i^* = 1,\ y_i = +1 \\ \log (1 - \pi_i) & y_i^* = 0,\ y_i = -1 \end{cases}$$

and $1 - \pi_i = 1 - \sigma(w^T x_i) = \sigma(-w^T x_i)$

$$E(w) = -\sum_{i=1}^{n} \begin{cases} \log \sigma(w^T x_i) & y_i^* = 1 & y_i = +1 \\ \log \sigma(-w^T x_i) & y_i^* = 0 & y_i = -1 \end{cases}$$

$$= -\sum_{i=1}^{n} \log \sigma(y_i w^T x_i)$$

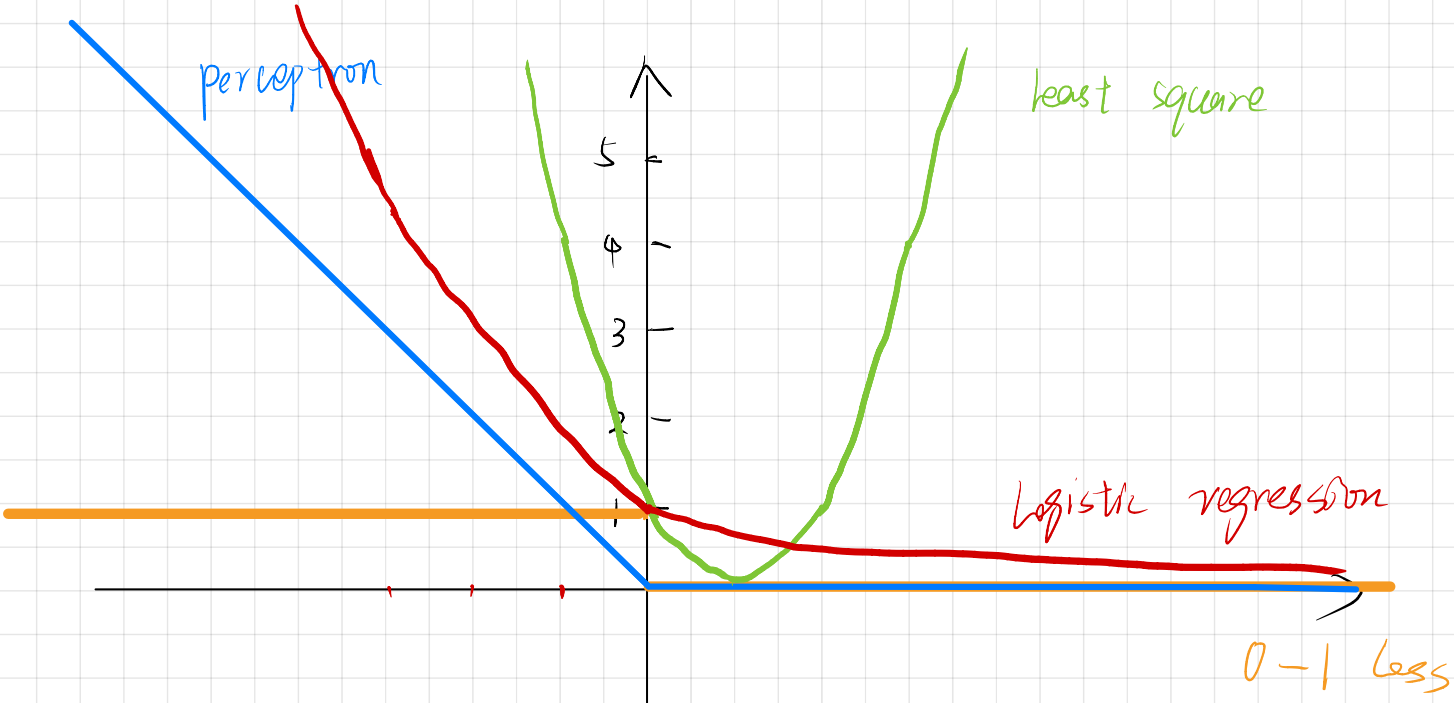
$$= -\sum_{i=1}^{n} \log \left[\frac{1}{1 + e^{-y_i w^T x_i}} \right]$$

$$= \sum_{i=1}^{n} \log (1 + e^{-y_i w^T x_i})$$

where $L_{LR}(z_i) = \log (1 + e^{-z_i}) \propto \frac{1}{\log 2} \log (1 + e^{-z_i})$

e)

In next page.



To linear classifier.

To 0-1 loss just have binary value for classifying

To perceptron it just focus on misclassify points.
but cannot make point away from the margin, And it depends on initialization.

To least square, it also give large loss when the predict is correct.

Above all, and logistic regression can give linear penalty, also can find the boundary margin for classes.
Thus, LR is better.

37558749
王[illegible]