

	Problem 2
-	· Find the result of minimum
-	· Find the result of minimum, the loss of sum of squared errors. · Add a pencity for an Lz penalty
	mr L1 Ornell
	$\frac{\text{argmin}}{w} \left\{ \sum \left(w^{T} = y_{i}^{2} \right)^{2} + \lambda \left\ w \right\ _{2}^{2} \right\}$
	. How does this change the solution to the original regression
	solution?
-	· What is the impact of adding the penalty?
-	
1	Logistic Function: It exp(-wx;)
	Cross Entropy: L(w) = - E{y; log(y;)+(1+y;) log(y;)}
	where ig = I+exp(-wix)
	where

argmin
$$\left\{ \sum_{i} \left(w^{T} x_{i} - y_{i}^{T} + \lambda \| w \|_{2}^{2} \right) \right\}$$

$$L(w) = \left(wx - y_{i}^{T} (wx - y_{i}) + \lambda w^{T} w \right)$$

$$= \overline{w} x^{T} \overline{w} x - \overline{w} x^{T} \overline{y} - \overline{y}^{T} \overline{w} x + \overline{y}^{T} \overline{y} + \lambda \overline{w}^{T} \overline{w}$$

$$= \overline{w} x^{T} \overline{w} x - \overline{w} x^{T} \overline{y} - \overline{y}^{T} \overline{w} x + \overline{y}^{T} \overline{y} + \lambda \overline{w}^{T} \overline{w}$$

$$= \overline{w} x^{T} \overline{w} x - 2 \overline{y}^{T} x \overline{w} + \overline{y}^{T} \overline{y} + \lambda \overline{w}^{T} \overline{w}$$

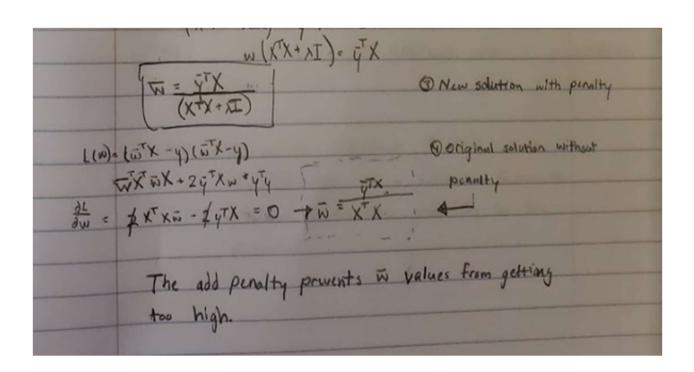
$$= \overline{w} x^{T} x^{T} \overline{w} x - 2 \overline{y}^{T} x \overline{w} + \overline{y}^{T} \overline{y} + \lambda \overline{w}^{T} \overline{w}$$

$$= \overline{w} (2 \overline{y} x^{T} w) = 2 \overline{y}^{T} x$$

$$= \overline{w} (x^{T} x \overline{w} - 2 \overline{y}^{T} x + x \overline{w}) = 0$$

$$= \overline{w} (x^{T} x + \lambda \overline{x}) - \overline{y}^{T} x = 0$$

$$= \overline{w} (x^{T} x + \lambda \overline{x}) - \overline{y}^{T} x = 0$$



- the Baye decision rule describes an opproach we take for choosing a so for X.
- · Can be athieved by modeling P(w|x), or $\frac{P(x|w)P(w)}{P(x)}$ · What are the Ref Cons for each approach?
- · why is know P(x) useful?

P(w/x): Directly estimate to probability data belongs to a class, however difficult to determine outliers.

P(x/w)P(w): Obtain useful information of the data by
P(x) finding the prior, likelihood, evidence, probabilities

P(X) : Since the likelihood P(X) can be

II, P(X) acts as a normalize Easter to

give on appropriate probability for the

posterior