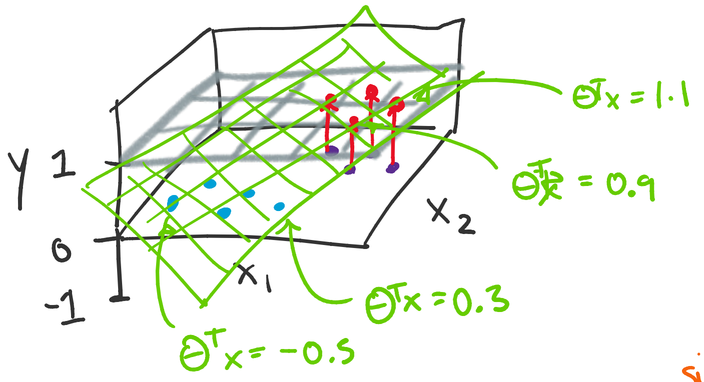
Section B: Linear Models + Feature Eng. + Regularization

Friday, September 30, 2022 1:26 PM

Why is it called Logistic Regression?

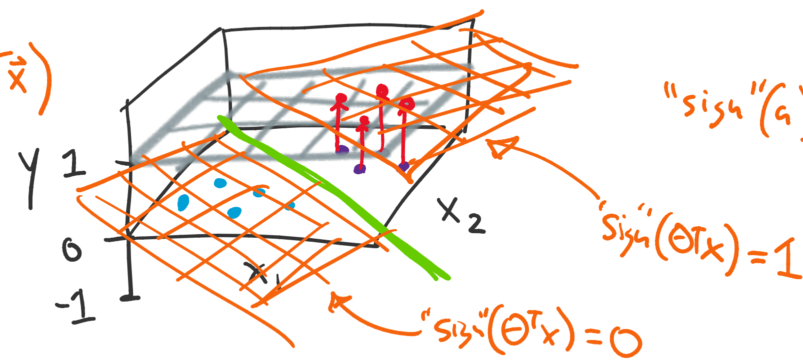
① $\vec{\theta}^T \vec{x}$



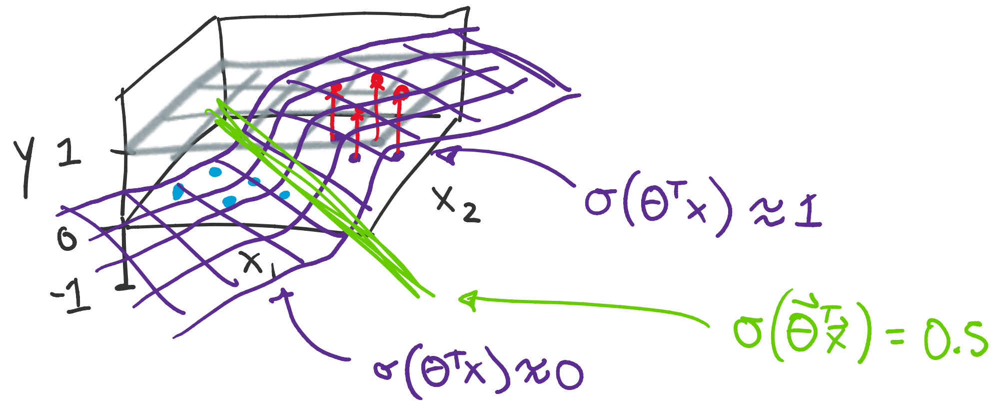
$$\text{sign}(a) = \begin{cases} +1 & \text{if } a \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

② "sign"$(\vec{\theta}^T \vec{x})$

$$\text{"sign"}(a) = \begin{cases} 1 & \text{if } a \geq 0 \\ 0 & \text{otherwise} \end{cases}$$



③ $\sigma(\vec{\theta}^T \vec{x})$



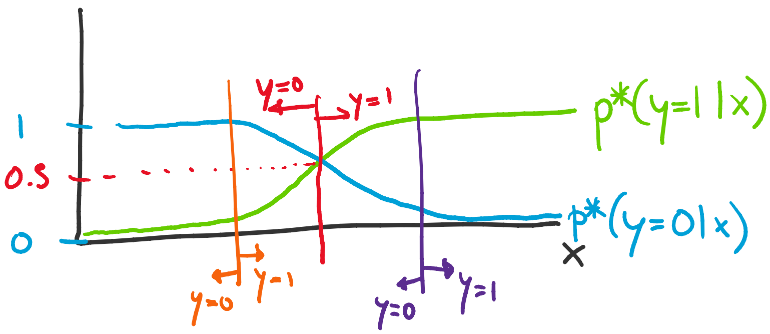
Why?

Log. Reg. is learning a function $\sigma(\vec{\theta}^T \vec{x})$ that regresses points $\vec{x}$ to outputs $y \in [0,1]$

BUT what we want is the decision boundary that it induces.

Bayes Optimal Classifier

Assume $M=1$ and we know $p^*(y|x)$



Loss function

$$\ell(\hat{y}, y^*) = \mathbb{I}(\hat{y} \neq y^*)$$

Average Loss

$$AL(\vec{x}, \hat{y}) = \sum_{y'} p^*(y'|\vec{x})\, \ell(\hat{y}, y')$$

Bayes Classifier

$$\boxed{h(\vec{x}) = \underset{\hat{y}}{\operatorname{argmin}}\; AL(x, \hat{y})}$$

$$= \underset{\hat{y}}{\operatorname{argmin}} \sum_{y'} p^*(y'|\vec{x})\, \mathbb{I}(\hat{y} \neq y')$$

$$= \underset{\hat{y}}{\operatorname{argmax}} - \sum_{y'} p^*(y'|\vec{x})\, \mathbb{I}(\hat{y} \neq y')$$

$$= \underset{\hat{y}}{\operatorname{argmax}} \sum_{y'} p^*(y'|\vec{x})\, \mathbb{I}(\hat{y} = y')$$

$$= \underset{\hat{y}}{\operatorname{argmax}}\; p^*(\hat{y}|\vec{x})$$

$$\mathbb{I}(\text{proposition}) = \begin{cases} 1 & \text{if prop is true} \\ 0 & \text{otherwise} \end{cases}$$

$$s(\hat{y}, y^*) = \begin{cases} 1 & \text{if } \hat{y} \neq y^* \\ 0 & \text{otherwise} \end{cases}$$

$$\ell(\hat{y}, y^*) = \begin{cases} 1\text{ million} & \text{if } \hat{y} \neq y^* \wedge y^* = 1 \\ 1000 & \text{if } \hat{y} \neq y^* \wedge y^* = 0 \\ 0 & \text{otherwise} \end{cases}$$