2/17/2020.

- Recap of Logistic Regression.
- Regularization in Logistic Regression.
- Linear Regression.

$$\sigma(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}}.$$

For feature $\bar{x}$, $(\bar{w}, t)$,

$$\text{Prob}\left(1 \mid \bar{x}, (\bar{w}, t)\right) = \boxed{\sigma(\bar{w}\cdot\bar{x} - t)} = \frac{\exp(\bar{w}\cdot\bar{x}-t)}{1+\exp(\bar{w}\cdot\bar{x}-t)}.$$

$$\text{Prob}\left(-1 \mid \bar{x}, (\bar{w}, t)\right) = \frac{1}{1+\exp(\bar{w}\cdot\bar{x}-t)}.$$

$\bar{w}\cdot\bar{x} - t$ is called logit function (effort).

$= \boxed{\bar{w}^1\cdot\bar{x}^1 + \dots + \bar{w}^d\cdot\bar{x}^d - t}$ (weights of attributes)

Logistic Regression. (ML principle?)

$$\boxed{\max_{(\bar{w}, t)} \prod_{i=1}^{n} \text{Prob}(y_i \mid \bar{x}_i, (\bar{w}, t)).}$$

$$\arg\max_{(\bar{w}, t)} \prod_{i=1}^{n} \text{Prob}(y_i \mid \bar{x}_i, (\bar{w}, t)).$$

$$\Leftrightarrow \arg\max_{(\bar{w},t)} \underbrace{J_s(\bar{w},t)}\left(= \sum_{i=1}^{n} \log \mathrm{Prob}(y_i \mid \bar{x}_i, (\bar{w},t))\right)$$

$J_s(\bar{w},t)$ is concave fn of $(\bar{w},t)$.

(at most one maxima) use gradient descent (ascent).

Let $\bar{x}^* \triangleq (\bar{x}^1, \ldots, \bar{x}^d, -1)$ (d+1 dimensional)

$$\bar{w}\cdot\bar{x} - t = (\bar{w},t)\cdot\bar{x}^*$$

$$\frac{\partial}{\partial t} J_s(\bar{w},t)$$

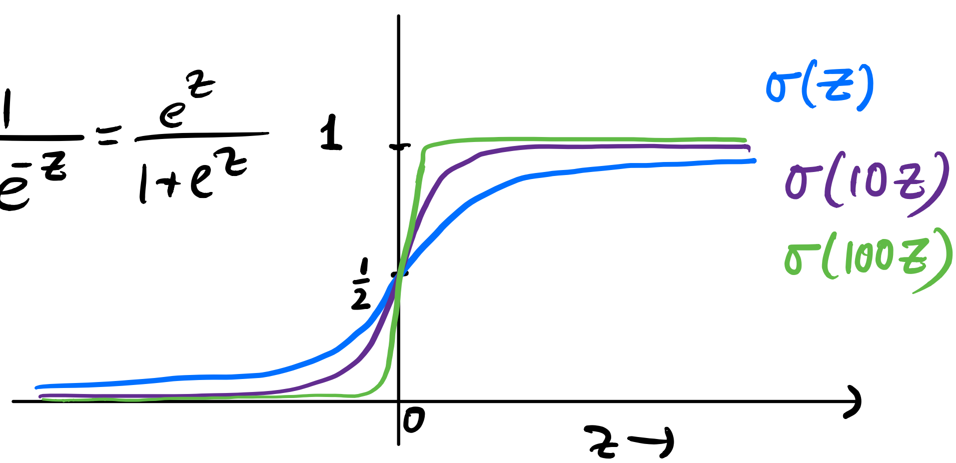
$$\nabla J_s(\bar{w},t) = \sum_{i=1}^{n} \left[\frac{1+y_i}{2} - P(1 \mid \bar{x}_i^*, \bar{w}, t)\right] \bar{x}_i^*$$

$$\boxed{(\bar{w}_j, t_j) \leftarrow (\bar{w}_{j-1}, t_{j-1}) + \eta \cdot \nabla J_s(\bar{w},t)}$$

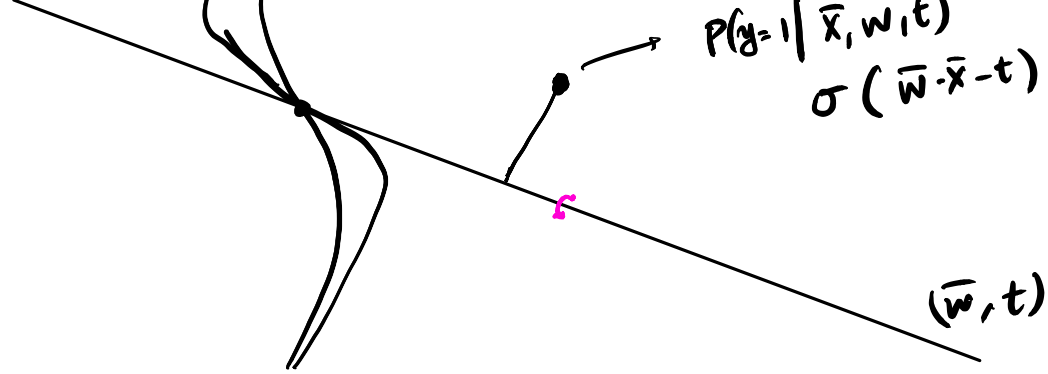
Regularization in logistic regression.

Discuss now, and then come back in linear regression.

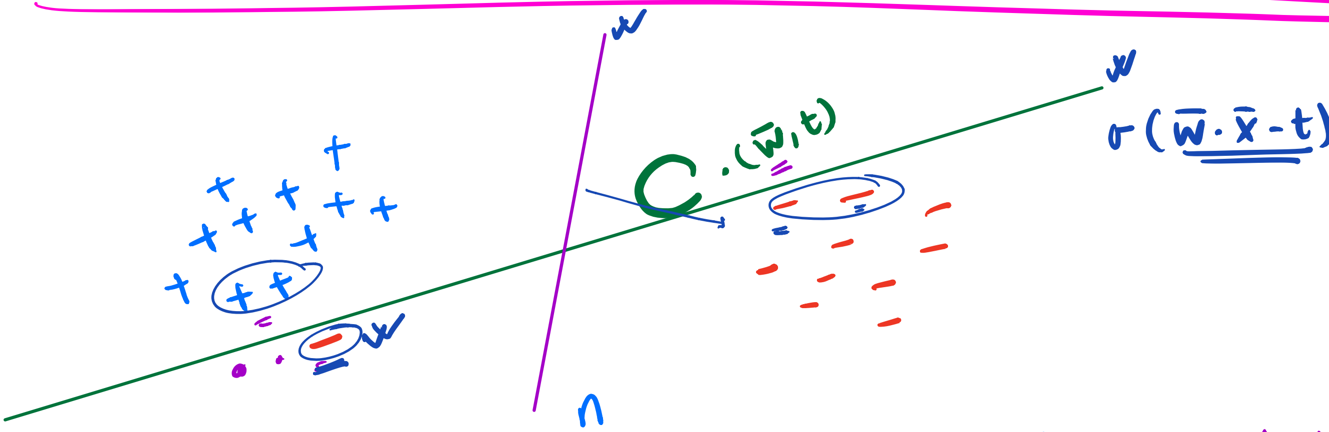
$$\sigma(z) = \frac{1}{1+e^{-z}} = \frac{e^z}{1+e^z}$$



$$(\bar{w},t) \longrightarrow (100\bar{w}, 100t)$$



A small change in $\bar{X}$ can make $P(1 \mid \bar{x}, \bar{w}, t)$ from 1 to 0



$$\arg \max_{(\bar{w}, t)} \prod_{i=1}^{n} \text{Prob}(y_i \mid \bar{x}_i, (\bar{w}, t)) \longrightarrow \text{'1'}$$

$$\longrightarrow (\bar{w}, t) - \boxed{\|\bar{w}\|_2 < C.}$$

Regularization :- Penalize models with large coefficients.

$$\arg \max_{(\bar{w}, t)} J_s^R(\bar{w}, t) = J_s(\bar{w}, t) - \frac{\lambda}{2} \|\bar{w}\|_2^2 .$$

$\lambda \rightarrow$ hyperparameter.

$$\boxed{\nabla = \nabla J_s(\bar{w}, t) - \lambda \cdot (\bar{w}, 0)}$$

* There will always be a unique maxima.

LINEAR REGRESSION

$S = \{(\bar{x}_1, y_1), (\bar{x}_2, y_2) \dots, (\bar{x}_n, y_n)\}.$

$y_i \in \mathbb{R}$, continuous/real valued.

y = price of a house

$\bar{x} = (\text{sqft}, \#\text{BR})$ $(d = 2)$

$\bar{x} = (\text{sqft})$ $(d = 1)$

Linear Models :- $y = \bar{w}^0 + \bar{w}^1 \cdot \bar{x}^1 + \bar{w}^2 \cdot \bar{x}^2 + \dots + \bar{w}^d \cdot \bar{x}^d$

$y = \bar{w}^0 + \bar{w}^1 \cdot (\text{sqft}) + \boxed{\bar{w}^2 \cdot (\#\text{BR})}.$ ×

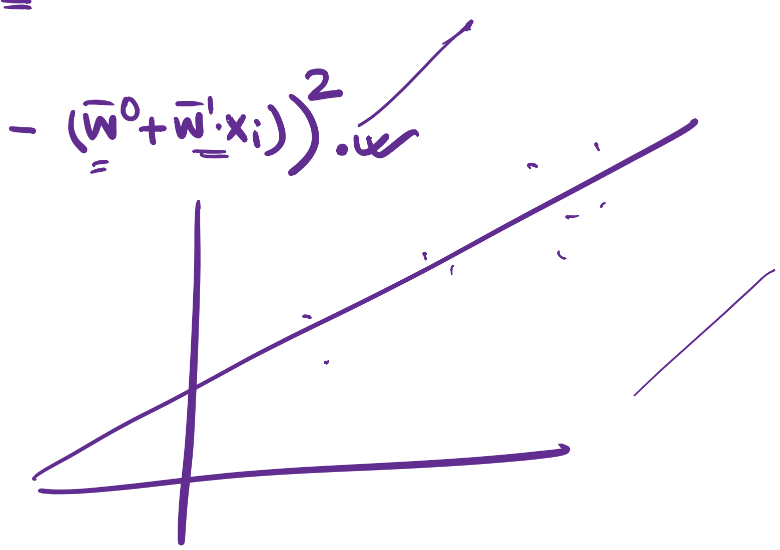
$y = \bar{w}^0 + \bar{w}^1 \cdot (\text{sqft})^2.$

Goal :- Find $(\bar{w}^0, \dots, \bar{w}^d)$ from 'S'.

Least squares Regression.

$d = 1$, $y = \bar{w}^0 + \bar{w}^1 \cdot \bar{x}^1 = \bar{w}^0 + \bar{w}^1 \cdot x.$ $x_i \in \mathbb{R}^1.$

$$\arg\min_{\bar{w}^0, \bar{w}^1} J(\bar{w}^0, \bar{w}^1) = \sum_{i=1}^{n} \left(y_i - (\bar{w}^0 + \bar{w}^1 \cdot x_i)\right)^2.$$



$\nabla J = 0.$

$$\frac{\partial J(\bar{w}^0, \bar{w}^1)}{\partial \bar{w}^0} = 0, \qquad \frac{\partial J(\bar{w}^0, \bar{w}^1)}{\partial \bar{w}^1} = 0.$$

$$\Rightarrow \sum_{i=1}^{n} 2(y_i - \bar{w}^0 - \bar{w}^1 x_i)\cdot(-1) = 0 \qquad \Rightarrow \sum_{i=1}^{n} 2\cdot(y_i - \bar{w}^0 - \bar{w}^1 x_i)\cdot(-x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} (y_i - \bar{w}^0 - \bar{w}^1 x_i) = 0 \quad - (a) \qquad \Rightarrow \sum_{i=1}^{n} x_i(y_i - \bar{w}^0 - \bar{w}^1 x_i) = 0. \quad - (b)$$

$$\Rightarrow \quad \bar{w}^0 = \frac{\sum y_i - \bar{w}^1(\sum x_i)}{n} \quad - (c)$$

...

$$\left.\begin{aligned} \bar{w}^1 &= \frac{n(\sum x_i y_i) - (\sum y_i)(\sum x_i)}{n\cdot(\sum x_i^2) - (\sum x_i)^2} \checkmark \\ \bar{w}^0 &= \frac{(\sum y_i)(\sum x_i^2) - (\sum x_i)(\sum y_i)}{n\cdot(\sum x_i^2) - (\sum x_i)^2}. \checkmark \end{aligned}\right\} \text{compute.}$$

For real numbers $a_1 \dots a_n \in \mathbb{R}$.

$$n\cdot(\sum a_i^2) \geq (\sum a_i)^2.$$

$$2\,(a_1^2 + a_2^2) \geq (a_1 + a_2)^2 \Leftrightarrow a_1^2 + a_2^2 \geq 2a_1a_2$$

$$\Leftrightarrow (a_1 - a_2)^2 \geq 0.$$
