Last class (10 Aug)

1) Regularization
2) L1 and L2 Regularization
3) Parameter vs Hyper-parameter
4) Cross-Validation
5) K-fold CV

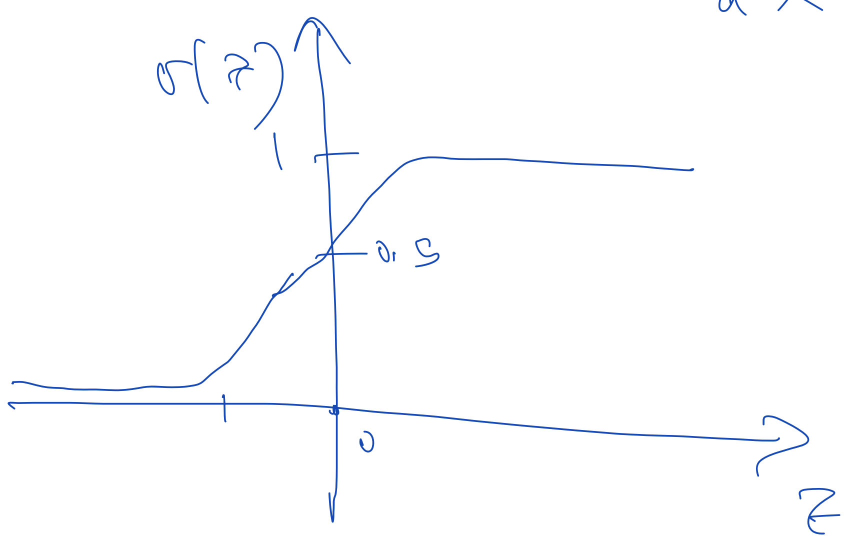
Today's class

1) Recap — Quizzes
2) AT&T Churn Prediction Problem
3) Recap of Linear Regression
4) Intro to Logistic Regression
5) Thresholding and Step Function
6) Sigmoid Function
7) Geometric Interpretation
8) Maximum likelihood
9) Logloss and optimization

$$x^{(i)} = [1 ; x_1 ; x_2 ; x_3 ; \dots x_d]$$

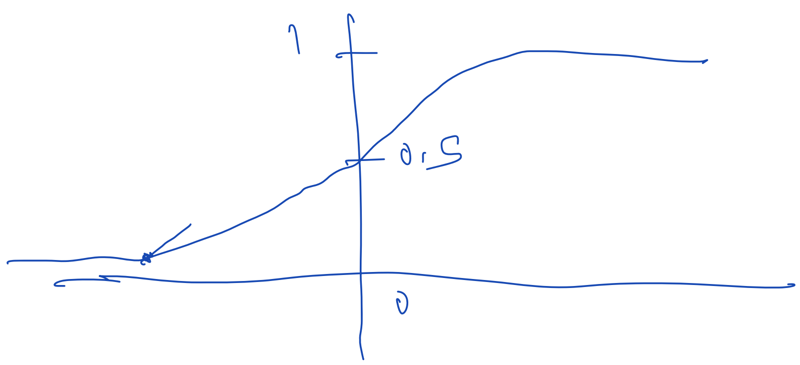
$$W = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix} \quad (d+1) \times 1$$

$$W x = [w_0 \times 1 + w_1 \times x_1 + w_2 \times x_2 + \dots w_d \times x_d]$$



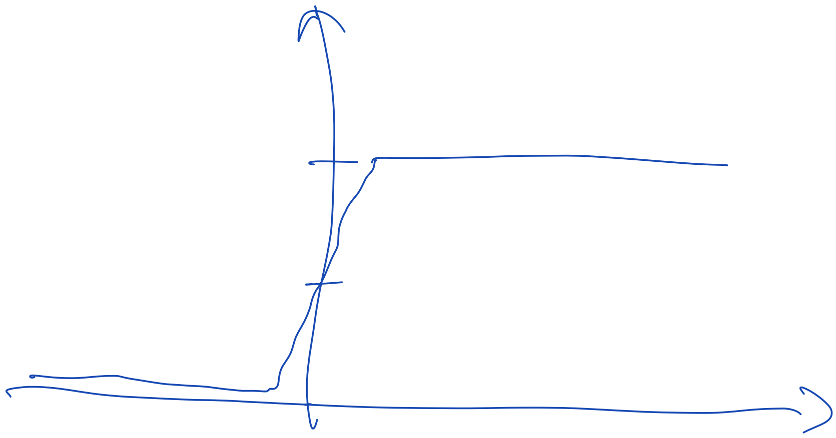
Sigmoid

$\sigma(z)$



$\sigma(z/K)$

$K<1, \ K>1$



$$\text{log-loss}\left(y^{(i)}, \hat{y}^{(i)}\right)$$

$$= -\left[y^{(i)} \log\left(\hat{y}^{(i)}\right) + \left(1 - y^{(i)}\right) \log\left(1 - \hat{y}^{(i)}\right) \right]$$

$y^{(i)} = 1$, $\hat{y}^{(i)} = 0.99$

close

$P\left(\hat{y}^{(i)} = 1 \mid x^{(i)}\right)$

error should be low

Lin Reg : MSE loss + Regularization Loss

$$\frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} - \hat{y}^{(i)}\right)^2 + \lambda_2 \sum_{j=1}^{d} w_j^2$$

$$+ \lambda_1 \sum_{j=1}^{d} |w_j|$$

Log Reg : Log. loss + Regul. loss

$$w^T x^{(i)} + w_0$$

$$= w_0 + w_1 x_1^{(i)} + w_2 x_2^{(i)} + \dots + \boxed{w_j x_j^{(i)}} + \dots w_d x_d^{(i)}$$

$$x_j^{(i)}$$

$$\frac{\partial\left(w^T x^{(i)} + w_0\right)}{\partial w_j} = x_j^{(i)}$$

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

$$\frac{d\,\sigma(z)}{dz} = \frac{d}{dz}\left(\frac{1}{1+e^{-z}}\right)$$

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

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

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

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

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

$$= \sigma(z) - \sigma(z)^2$$

$$= \sigma(z)\left[1 - \sigma(z)\right]$$