Linear Regression:

Input: $x \in \mathbb{R}^{d+1}$

Ouput: $y \in \mathbb{R}$

Parameters: $\theta \in \mathbb{R}^{d+1}$

Model: $h(x) = \theta^T x$

Loss function: $J(\theta) = \frac{1}{2} \sum_{i=1}^{n} (\theta^T x^{(i)} - y^{(i)})^2$

Optimization: $\theta^* = (X^T X)^{-1} X^T y$

$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$

$y = \theta_1 x + \theta_0$

$\downarrow$

$y = \theta_1 x + \theta_2 x_2 + \cdots + \theta_d x_d + \theta_0$

$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_d \end{bmatrix} \quad x = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_d \end{bmatrix}$

Probabilistic (Maximum Likelihood)

$y = \theta_1 x + \theta_0 + \varepsilon \rightarrow$ noise $\quad \varepsilon \sim N(0, \sigma^2)$

$N(\mu, \sigma) = \frac{1}{\sqrt{2\pi}} e$

$P(\varepsilon^{(i)}) \sim N(0, \sigma^2)$, thus $P(y^{(i)} | x^{(i)}, \theta)$

$L(\theta) = \prod_{i=1}^{n} P(y^{(i)} | x^{(i)}, \theta)$.

$\theta_{ML} = \underset{\theta}{\operatorname{argmax}} \log L(\theta) = \sum_{i=1}^{n} P(y^{(i)} | x^{(i)}, \theta)$

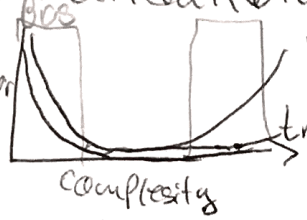
Squared loss $J(\theta) = \frac{1}{2} \sum_{i=1}^{n} (h(x^{(i)}) - y^{(i)})^2 = \frac{1}{2} (X\theta - y)^T (X\theta - y)$

$\nabla_\theta J(\theta) = \nabla_\theta \left(\frac{1}{2} y^T y - \theta^T X^T y + \frac{1}{2} \theta^T X^T X \theta \right) = -X^T y + X^T X \theta$

$0 = -X^T y + X^T X \theta$

$\theta^* = (X^T X)^{-1} X^T y$

- Underffitting vs overfitting: Polynomial regression
- Generalization Error: training set, test set
- Validation set is an extra set to validate the model and we dont touch the test set until the end.
- Cross-validation (K-fold validation): avoid wasting data in validation-set
- Bias-variance trade-off:



Bias & variance: $\hat{\theta}$ is an stimator for θ: $bias(\hat{\theta}) = E[\hat{\theta}] - \theta$

$Var(\hat{\theta}) = E[(\hat{\theta} - E[\theta])^2]$, $E[(\hat{\theta} - \theta)^2] = bias(\hat{\theta})^2 + Var(\hat{\theta})$

Ridge Regression: $J(\theta) + \frac{\lambda}{2}\|\theta\|_2^2, \lambda > 0$

Solution: $\theta^* = (X^T X + \lambda I)^{-1} X^T y$

(2,4)

$3x_1 + 5x_2 - 15 \geq 0$

$3(3) + 5(5) - 15$

$3(6) + 5(2) - 15 \geq 0$

$0 = 3x_1 + 5x_2 - 15$

$3x_1 + 5x_2 = 15$

$\frac{3x_1}{5} = \frac{5x_2}{5} + \frac{15}{5}$

Logistic Regression (Classification) (yes or no)

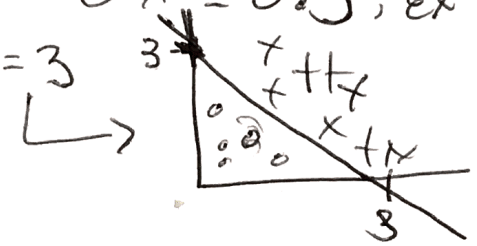
- sigmoid maps the whole real axis into finite intervals

$\sigma(t) = \frac{1}{1+e^{-t}}$, $P(y=1 | x; \theta) = h(x) = \sigma(\theta^T x) = \frac{1}{1-e^{-\theta^T x}}$

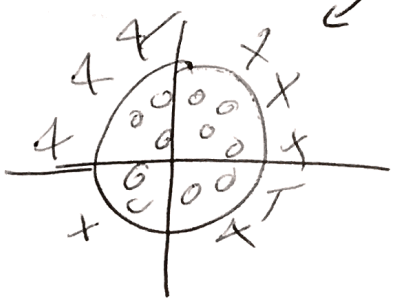
Decision Boundary

$y=1$ if $h(x) \geq 0.5 \Rightarrow \theta^T x \geq 0.5$, ex: $h(x) = \sigma(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$

and $\theta = (3,1,1)$ $x_1 + x_2 = 3$



$x_1^2 + x_2^2 = 1$



Nonlinear decision boundaries

$h(x) = \sigma(\theta_0 + x_1\theta_1 + x_2\theta_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$

$\theta(-1,0,0,1,1)$

Cost function $J(\theta) = \frac{1}{n}\sum_{i=1}^{n} \text{cost}(h(x)^{(i)}, y^{(i)})$

$3x_1 + 5x_2 \geq 15$

$x_1 \geq \frac{5}{3}x_2 + 5$

$x_2 \geq \frac{3}{5}x_1 + 3$

$\text{Cost}(h(x), y) = -y \log(h(x)) - (1-y)\log(1-h(x))$

thus $J(\theta) = -\frac{1}{n}\sum_{i=1}^{n} [y^{(i)} \log(h(x^{(i)})) + (1-y^{(i)}) \log(1-h(x^{(i)}))$

MLE: $P(y=1 | x, \theta) = \sigma(\theta^T x)$, $P(y=0 | x, \theta) = 1 - \sigma(\theta^T x)$

$L(\theta) = \prod_{i=1}^{n} \sigma(\theta^T x^{(i)})^{y^{(i)}} (1 - \sigma(\theta^T x^{(i)}))^{1-y^{(i)}}$

$-\log(L(\theta)) = -\sum_{i=1}^{n} [y^{(i)} \log(\sigma(\theta^T x^{(i)})) + (1-y^{(i)}) \log(1 - \sigma(\theta^T x^{(i)}))$