# Lecture 5:
# Linear Models Review

**Rui Xia**
School of Computer Science and Engineering
Nanjing University of Science and Technology

# Linear Models (We Learnt  So Far)

- Linear Regression
- Logistic Regression
- Perceptron Algorithm

# 3 Key Concepts in Machine Learning

- Hypothesis
  - Math models with (unknown) parameters (or structures)
- Learning (to estimate the parameters)
  - Maximum Likelihood Estimation (MLE), MAP, Bayesian Estimation
  - Cost Function Optimization
- Decision
  - Bayes decision rule
  - Direct prediction function

# Model Hypothesis

- Linear Regression

$$h_\theta(x) = \theta^T x$$

- Perceptron Algorithm

$$h_\theta(x) = \begin{cases} 1 & if \;\; \theta^T x \;\geq 0 \\ 0 & if \;\; \theta^T x \;< 0 \end{cases}$$

- Logistic Regression

$$h_\theta(x) = \delta(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$

$$P(y = 1|x; \theta) = h_\theta(x) \qquad\qquad P(y = 0|x; \theta) = 1 - h_\theta(x)$$

# Learning Criteria (Cost Functions)

- Linear Regression

$$J_l(\theta) = \frac{1}{2}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

**Maximum Likelihood ⇔ Least Mean Square**

# Learning Criteria (Cost Functions)

- Perceptron Algorithm

$$J_p(\theta) = \sum_{x^{(i)} \in M_0} \theta^T x^{(i)} - \sum_{x^{(j)} \in M_1} \theta^T x^{(j)}$$

$$= \sum_{i=1}^{m} \left( \left(1 - y^{(i)}\right) h_\theta\left(x^{(i)}\right) - y^{(i)} \left(1 - h_\theta\left(x^{(i)}\right)\right) \right) \theta^T x^{(i)}$$

$$= \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \theta^T x^{(i)}$$

**Perceptron Criterion**

# Learning Criteria (Cost Functions)

- Logistic Regression

$$J_c(\theta) = \sum_{i=1}^{m} y^{(i)} log h_\theta\left(x^{(i)}\right) + \left(1 - y^{(i)}\right) log \left(1 - h_\theta\left(x^{(i)}\right)\right)$$

**Maximum Likelihood ⇔**
**Minimum Cross Entropy Error**

# Gradient Descent Optimization

- Linear Regression

$$\frac{\partial}{\partial\theta}J_l(\theta) = \frac{1}{2}\frac{\partial}{\partial\theta}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$
$$= \sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)x^{(i)}$$

$$\theta := \theta - \alpha\frac{\partial}{\partial\theta}J_l(\theta) = \theta - \alpha\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)x^{(i)}$$

# (Stochastic) Gradient Descent Optimization

- Perceptron Algorithm

$$\frac{\partial}{\partial\theta}J_p(\theta) = \sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)x^{(i)}$$

$$\omega := \omega + \alpha\left(y - h_\theta(x)\right)x$$

$$= \begin{cases} \omega + \alpha x & if\ y = 1\ and\ h_\theta(x) = 0 \\ \omega - \alpha x & if\ y = 0\ and\ h_\theta(x) = 1 \\ \omega & others \end{cases}$$

# Gradient Descent Optimization

- Logistic Regression

$$\frac{\partial J_c(\theta)}{\partial \theta} = \sum_{i=1}^{m} \left( y^{(i)} \frac{1}{h_\theta(x^{(i)})} - \left(1 - y^{(i)}\right) \frac{1}{1 - h_\theta(x^{(i)})} \right) \frac{\partial}{\partial \theta} h_\theta\left(x^{(i)}\right)$$

$$= \sum_{i=1}^{m} \left( y^{(i)} \frac{1}{h_\theta(x^{(i)})} - \left(1 - y^{(i)}\right) \frac{1}{1 - h_\theta(x^{(i)})} \right) h_\theta\left(x^{(i)}\right) \left(1 - h_\theta\left(x^{(i)}\right)\right) \frac{\partial}{\partial \theta} \theta^T x^{(i)}$$

$$= \sum_{i=1}^{m} \left( y^{(i)} \left(1 - h_\theta\left(x^{(i)}\right)\right) - \left(1 - y^{(i)}\right) h_\theta\left(x^{(i)}\right) \right) x^{(i)}$$

$$= \sum_{i=1}^{m} \left( y - h_\theta\left(x^{(i)}\right) \right) x^{(i)}$$

$$\theta := \theta + \alpha \sum_{i=1}^{m} \left( y^{(i)} - h_\theta\left(x^{(i)}\right) \right) x^{(i)}$$


# Any Questions?