Logistic Regression II

CMSC 422

SOHEIL FEIZI

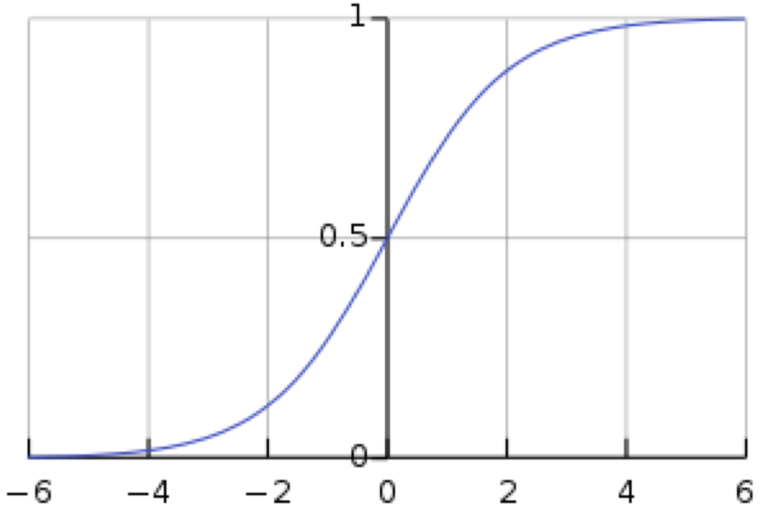
sfeizi@cs.umd.edu

Logistic Regression

- Binary classification

$$P(Y^{(i)} = 1|X^{(i)}, \theta) = g(< \theta, X^{(i)} >)$$

$$P(Y^{(i)} = 0|X^{(i)}, \theta) = 1 - g(< \theta, X^{(i)} >)$$



Sigmoid function

$$g(z) = \frac{1}{1 + \exp(-z)}$$

Logistic Regression

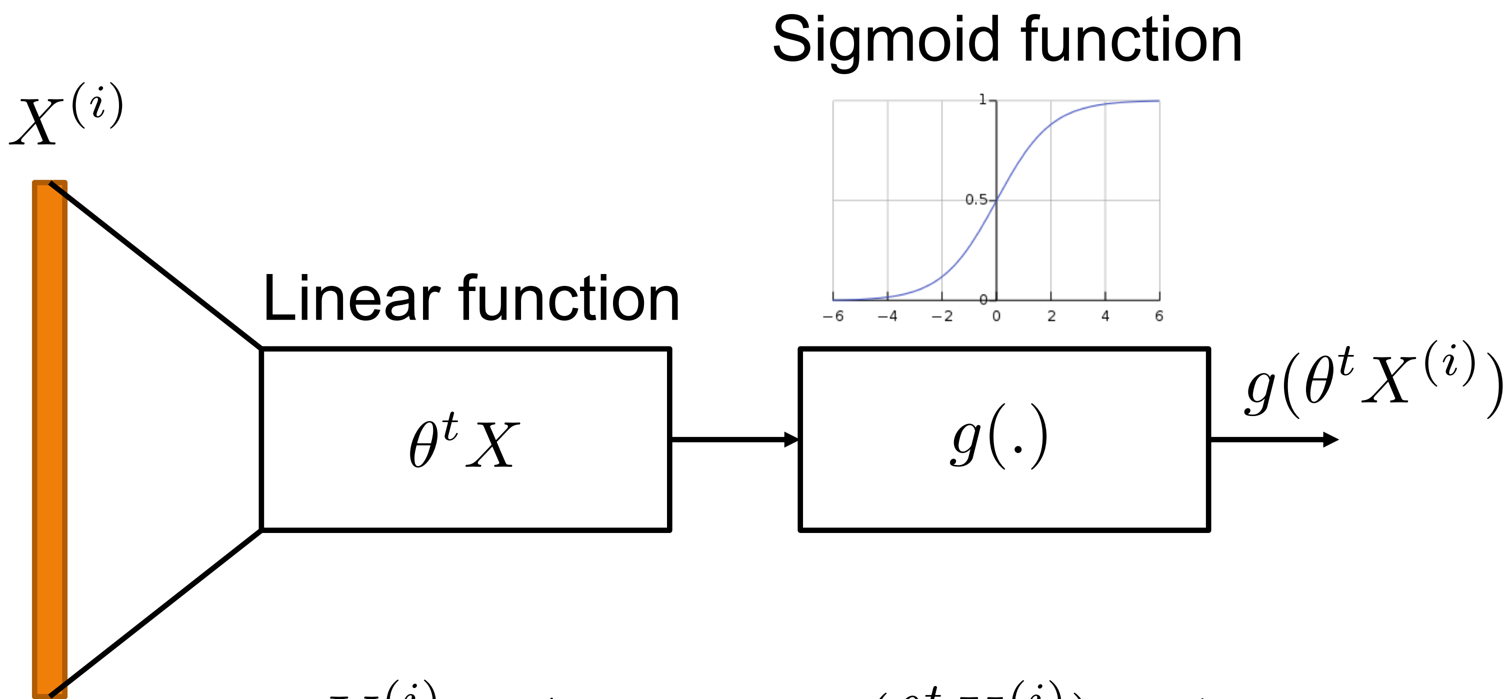
- Maximum Likelihood

$$\max_{\theta} \quad \underbrace{\sum_{i=1}^{N} Y^{(i)} \log g(< \theta, X^{(i)} >) + (1 - Y^{(i)}) \log(1 - g(< \theta, X^{(i)} >))}_{\text{Cross-entropy loss function}}$$

Cross-entropy loss function

- Convex optimization, solve using gradient descent

A High-Level View



$$Y^{(i)} = 1 \longrightarrow g(\theta^t X^{(i)}) \approx 1$$

$$Y^{(i)} = 0 \longrightarrow g(\theta^t X^{(i)}) \approx 0$$

Does cross entropy optimization encourage this?

Multi-Label Classification

- Suppose we have labels {0,1,…,k}

- How can we extend logistic regression's formulation for the general case?

Recall the probabilistic model

- In binary classification, we have

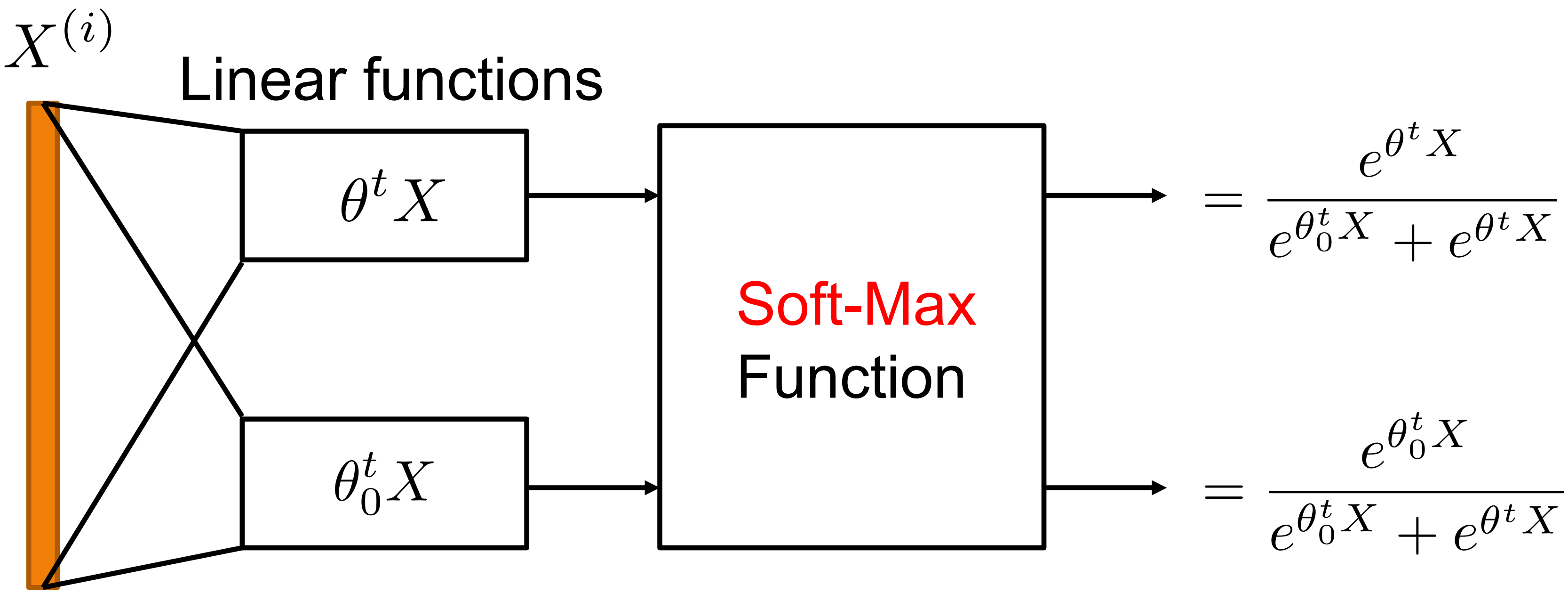
$$P(Y=1|X,\theta) = g(\theta^t X) \quad = \frac{1}{1+e^{-\theta^t X}} = \frac{e^{\theta^t X}}{1+e^{\theta^t X}}$$

$$= \boxed{\frac{e^{\theta^t X}}{e^{\theta_0^t X} + e^{\theta^t X}}}$$

$$P(Y=0|X,\theta) = 1 - g(\theta^t X) = \frac{1}{1+e^{\theta^t X}} = \boxed{\frac{e^{\theta_0^t X}}{e^{\theta_0^t X} + e^{\theta^t X}}}$$

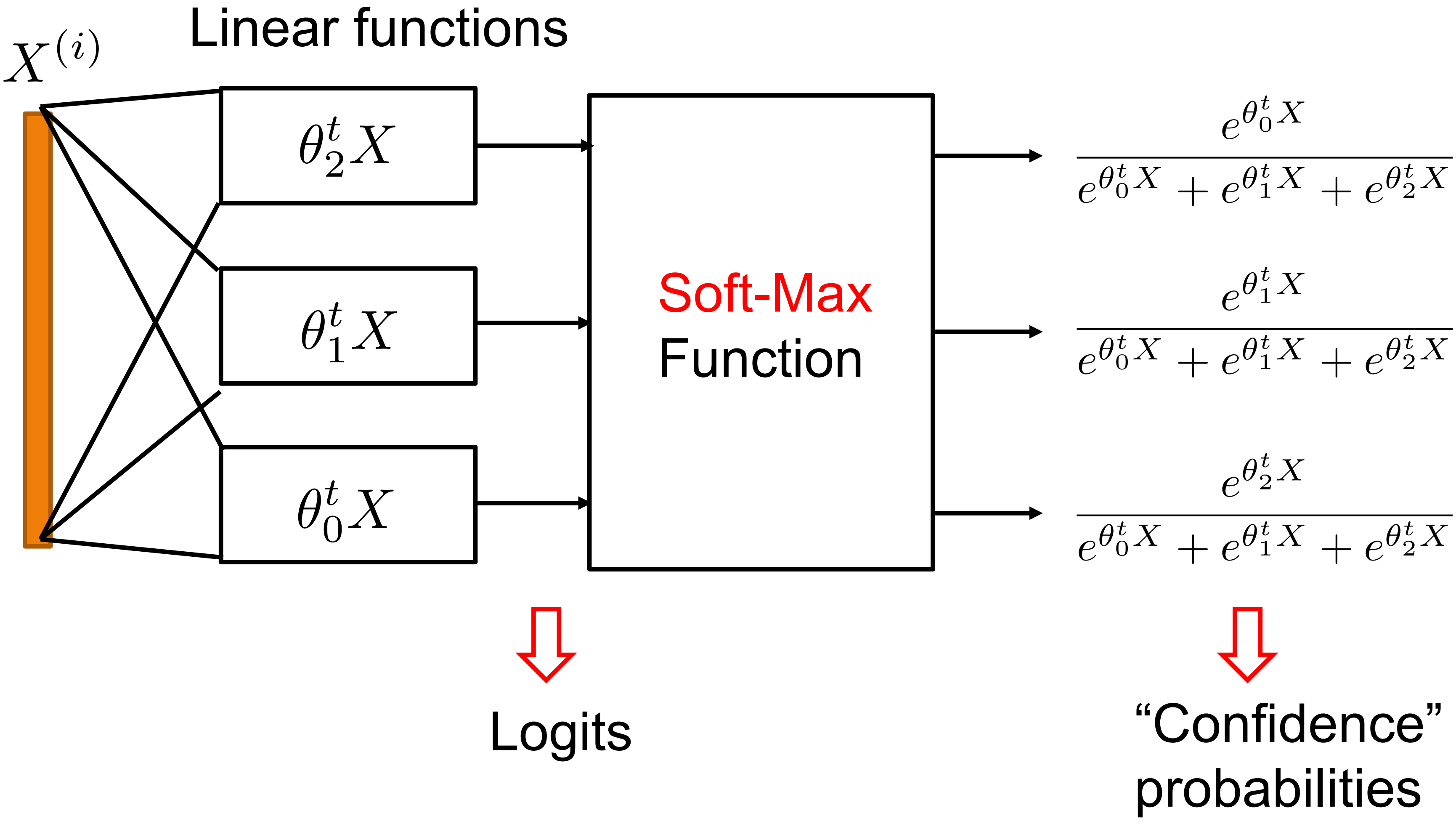
- If $\theta_0 = 0 \longrightarrow e^{\theta_0^t X} = 1$

A High-Level View:Binary Classification



How to extend this to the multi label classification?

Multi-Label Classification



Cross-Entropy Loss for Multi-Label Case

- Recall the binary case

$$\max_{\theta} \quad \sum_{i=1}^{N} Y^{(i)} \log g(< \theta, X^{(i)} >) + (1 - Y^{(i)}) \log(1 - g(< \theta, X^{(i)} >))$$

- Multi-label case

$$\sum_{\text{all samples}} 1\{Y^{(i)} = \text{label}\} \log(\text{corresponding confidence prob.})$$