

# Logistic Regression II

CMSC 422

SOHEIL FEIZI

[sfeizi@cs.umd.edu](mailto:sfeizi@cs.umd.edu)

# Today's topics

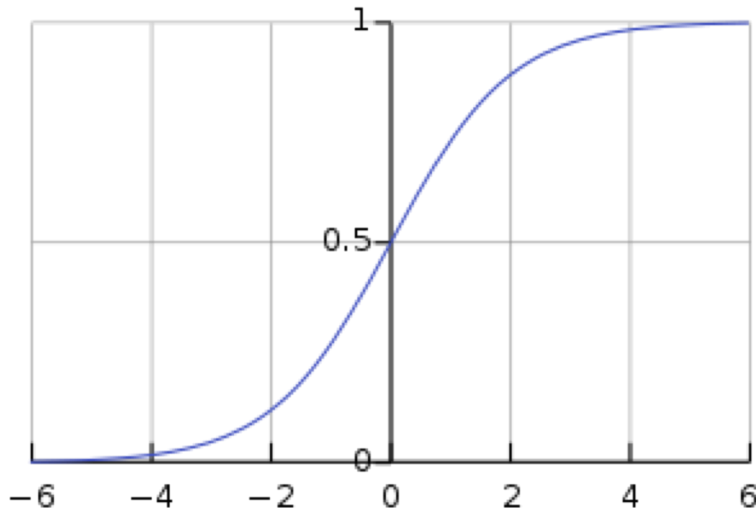
- Continue Discussion on Logistic Regression
- Multilabel Classifications
- HW4 deadline extended
- Final project topics/group members (deadline April 2nd)

# Logistic Regression

- Binary classification

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

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



Sigmoid function

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

# Logistic Regression

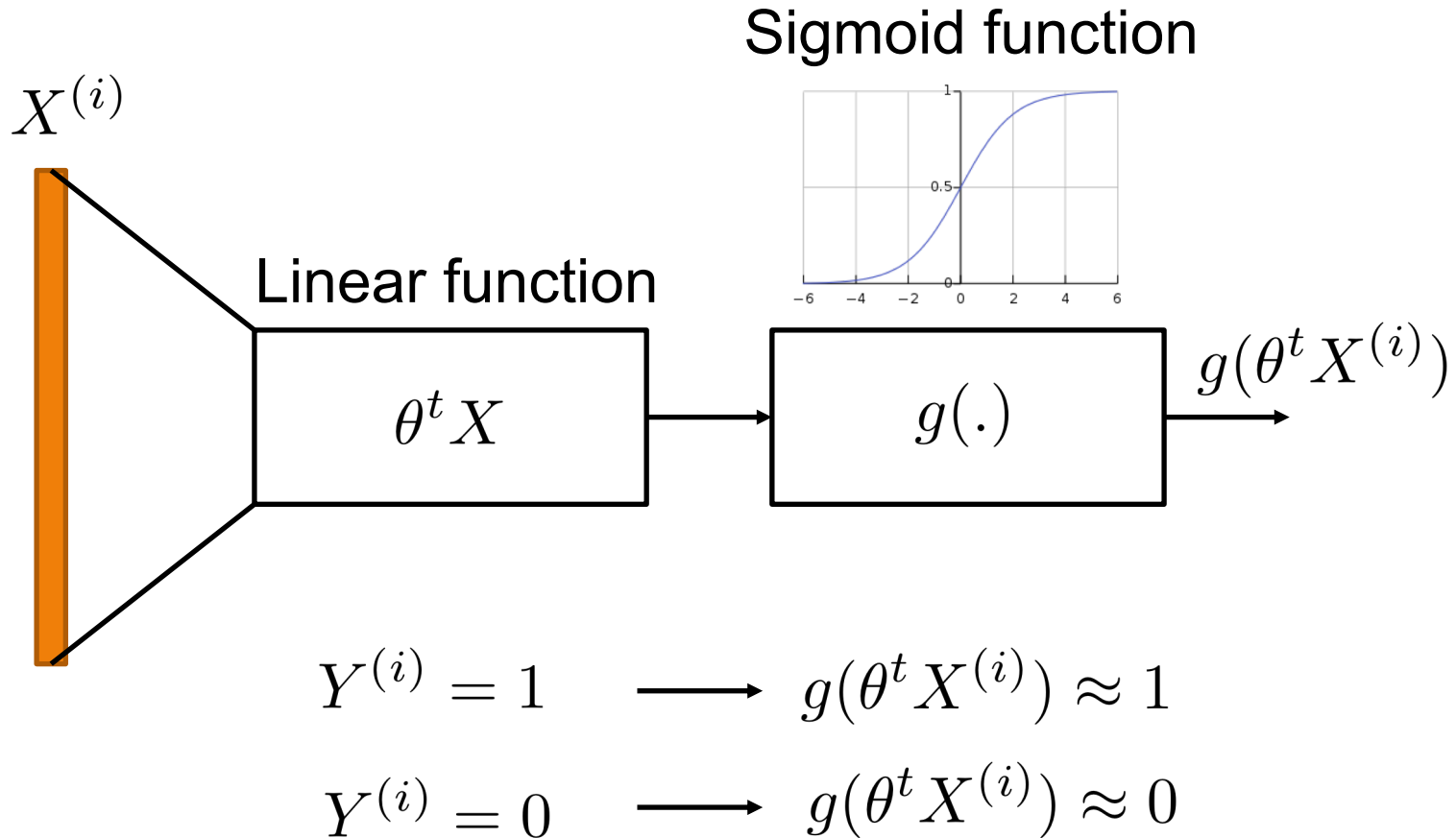
- Maximum Likelihood

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

**Cross-entropy** loss function

- Convex optimization, solve using gradient descent

# A High-Level View



Does cross entropy optimization encourage this?

# Multi-Label Classification

- Suppose we have labels  $\{0,1,\dots,k\}$
- How can we extend logistic regression's formulation for the general case?

# Recall the probabilistic model

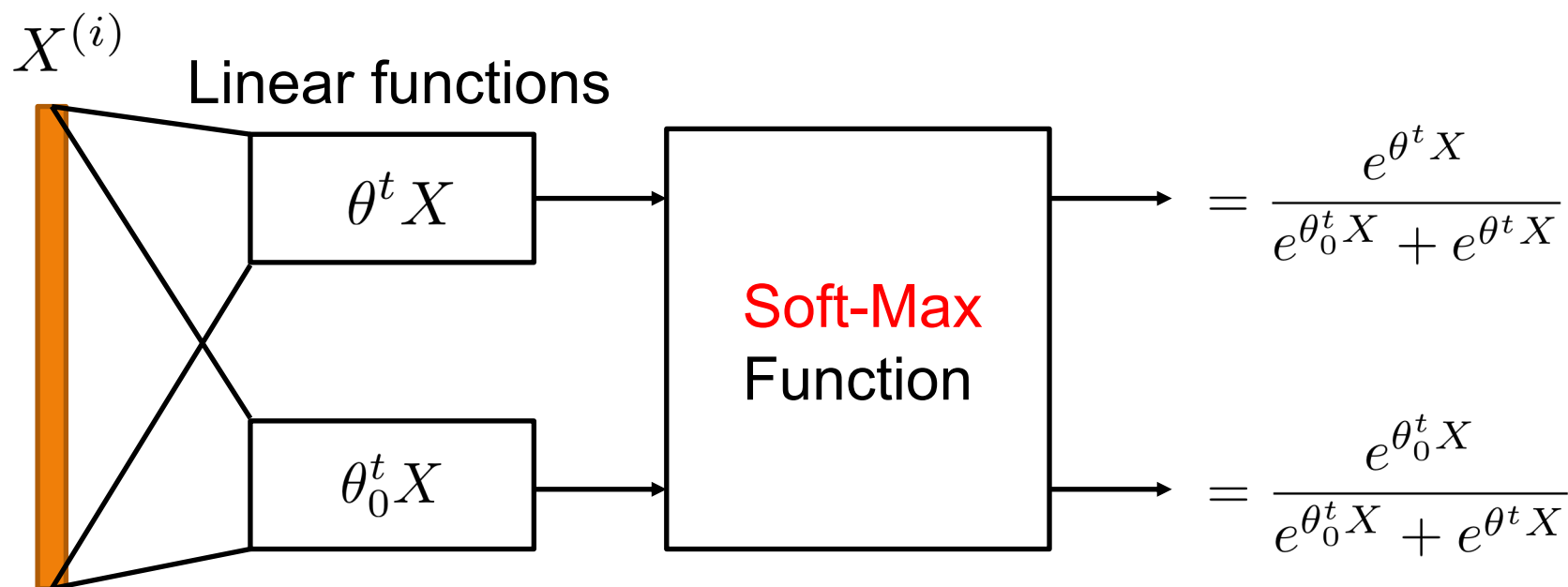
- In binary classification, we have

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

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

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

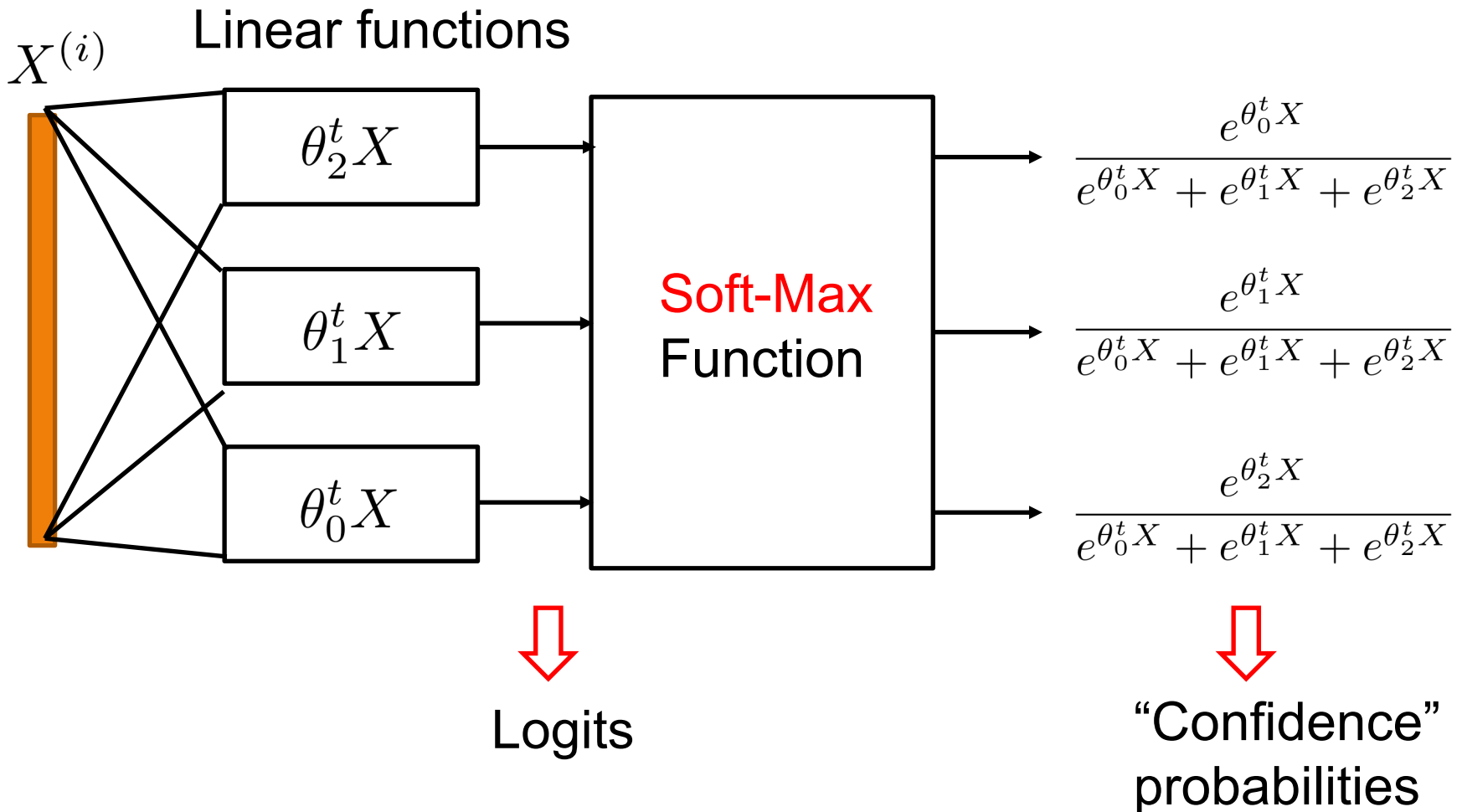
# A High-Level View: Binary Classification



How to extend this to the multi label classification?



# Multi-Label Classification



# Cross-Entropy Loss for Multi-Label Case

- Recall the binary case

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

- Multi-label case

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