

# Lecture 13 (Classification Problem)

## 1 Classification

1. Given  $\{x_i, y_i\}_{i=1}^m$  where  $y_i$  is discrete and *currently* takes only values 0, 1
2. The equation of separator is given as  $\theta^T x = 0$
3. The distribution is Bernoulli, i.e.,  $P(y_i = 1|x_i; \theta) = \Phi(\theta^T x_i)$  ( $\theta^T x_i$  gives the normal distance of  $x_i$  from the separator)
4.  $\Phi(z) = \frac{1}{1 + e^{-z}}$
5. Log likelihood ( $LL(\theta)$ ):

$$\log\left(\prod_{i=1}^m P(y_i \text{ is predicted correctly} | x_i; \theta)\right) = \sum_{i=1}^m \left( (y_i = 1) \log(\Phi(\theta^T x_i)) + (y_i = 0) \log(1 - \Phi(\theta^T x_i)) \right)$$

6.  $\theta_{ML} = \underset{\theta}{\operatorname{argmax}}(LL(\theta))$
7. This can now be solved using gradient descent, and  $\nabla_{\theta}(LL(\theta)) = \sum_{i=1}^m \left( y_i - \frac{1}{1 + e^{-\theta^T x_i}} \right) x_i$