Properties of the logistic function

If $p(y=0|x) = \frac{1}{1+e^{w^T x}}$ what function is $p(y=1|x)$?

$$p(y=1|x) = 1 - \frac{1}{1+e^{w^T x}} = \frac{e^{w^T x}}{1+e^{w^T x}}$$

What happens at $p(y=1|x) = p(y=0|x)$?

Properties of the logistic function

$$p(y=0|x) = p(y=1|x)$$

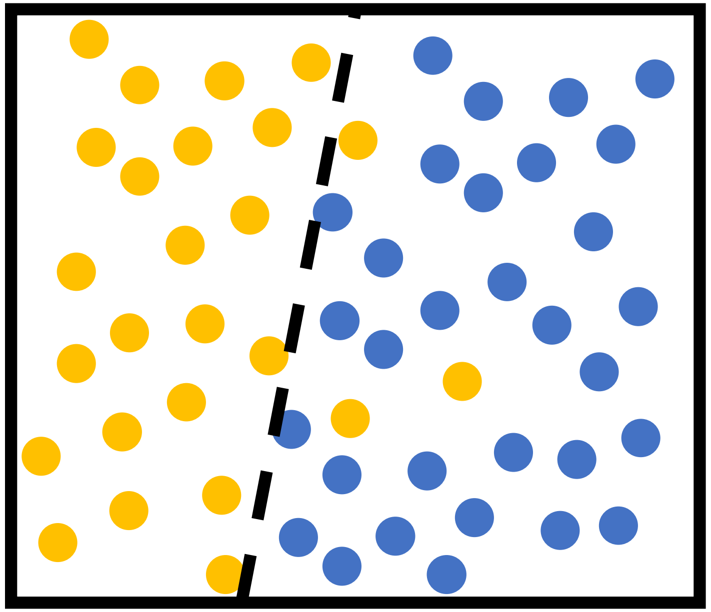
$$\frac{p(y=1|x)}{p(y=0|x)} = 1 = \frac{\frac{e^{w^T x}}{1+e^{w^T x}}}{\frac{1}{1+e^{w^T x}}} = e^{w^T x}$$

$$\log\left(\frac{p(y=1|x)}{p(y=0|x)}\right) = w^T x = 0$$

Log of the odds, or logit

Classification (probability) **Regression**

$$[0,1] \rightarrow [-\infty, +\infty]$$



Direct interpretation of the coefficients

$$\log\left(\frac{p(y=1|x)}{p(y=0|x)}\right) = w^T x = 0$$

Log of the odds, or logit

Keeping all the other quantities fixed, changing $\boldsymbol{x_i}$ by one unit, changes the logit by $\boldsymbol{w_i}$

How do we write an associated loss?

Two ways:

1) Derive it from a Maximum Likelihood approach.
2) Difference between true and estimated probability distributions.

Maximum Likelihood

- Likelihood: Product of event probabilities.
- For simplicity, let's take $p(y=1|x_i) = p(x_i)$ *($p(y=1|x_i)$ is $p(y=1|x)$ evaluated at the point x_i)*
- I can write the likelihood as:

$$\mathcal{L} = \prod_{i:y_i=1} p(x_i) \prod_{i:y_i=0} \left(1 - p(x_i)\right)$$

$$\Big\Downarrow \text{ Using } a^0 = 1$$

$$\mathcal{L} = \prod_i p(x_i)^{y_i} \left(1 - p(x_i)\right)^{1-y_i}$$

Maximum Likelihood

$$\mathcal{L} = \prod_i p(x_i)^{y_i} \left(1 - p(x_i)\right)^{1-y_i}$$

$\Big\Downarrow$ Taking logarithm

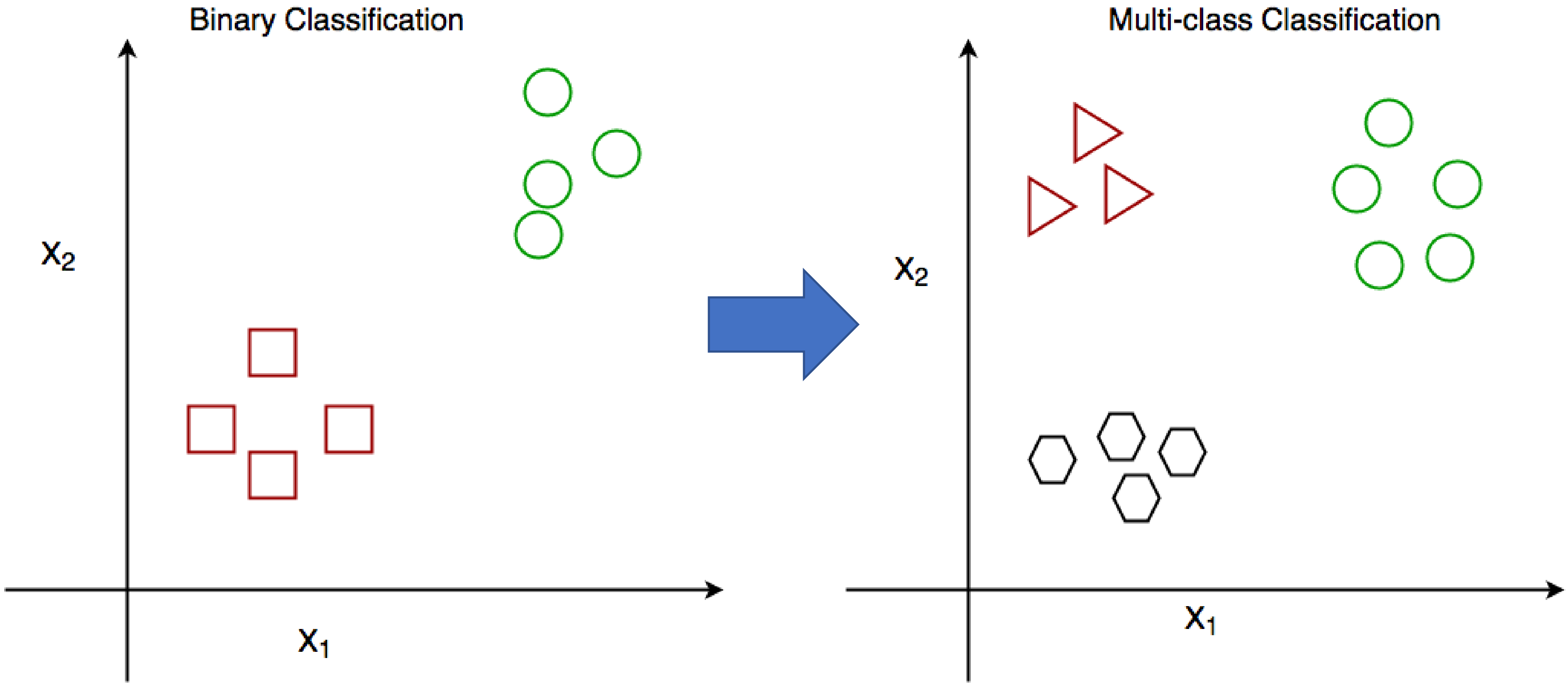
$$\mathcal{L} = \sum_i y_i \log p(x_i) + (1 - y_i) \log\left(1 - p(x_i)\right)$$

How would you transform a maximization problem in a minimization one?

Gradient descent of the logistic regression

$$w_{t+1} = w_t - \gamma \frac{\partial \mathcal{L}}{\partial w} = w_t - \gamma \left(\frac{1}{N} \sum_i x_i \left(\frac{1}{1 + e^{-w^T x_i}} - y_i \right) \right)$$

Extending the logistic model: Multinomial regression to K classes



Extending the logistic model: Multinomial regression to K classes

$$p(y=0|x) = \frac{1}{1+e^{w^T x}}$$

$$p(y=1|x) = \frac{e^{w^T x}}{1+e^{w^T x}}$$

$$p(y=1|x) = \frac{e^{w_1^T x}}{1+\sum_j^K e^{w_j^T x}}$$

$$p(y=2|x) = \frac{e^{w_2^T x}}{1+\sum_j^K e^{w_j^T x}}$$

…

$$p(y=K-1|x) = \frac{e^{w_{K-1}^T x}}{1+\sum_j^K e^{w_j^T x}}$$

$$p(y=K|x) = \frac{1}{1+\sum_j^K e^{w_j^T x}}$$

Extending the logistic model: Multinomial regression to K classes

$$\log\left(\frac{p(y=1|x)}{p(y=0|x)}\right) = w^T x \quad \Longrightarrow \quad \log\left(\frac{p(y=j|x)}{p(y=K|x)}\right) = w_j^T x$$

$$\mathcal{L} = -\sum_{j=1}^{M}\sum_{i}\left(\delta_{j,y_i}\log\left(p_j(x_i)\right) + \left(1-\delta_{j,y_i}\right)\log\left(1-p_j(x_i)\right)\right)$$

What happens with this loss?

if $j = y_i$: $\delta_{j,y_i} = 1$; else $\delta_{j,y_i} = 0$

$$p_j(x_i) = p(y = j|x_i)$$