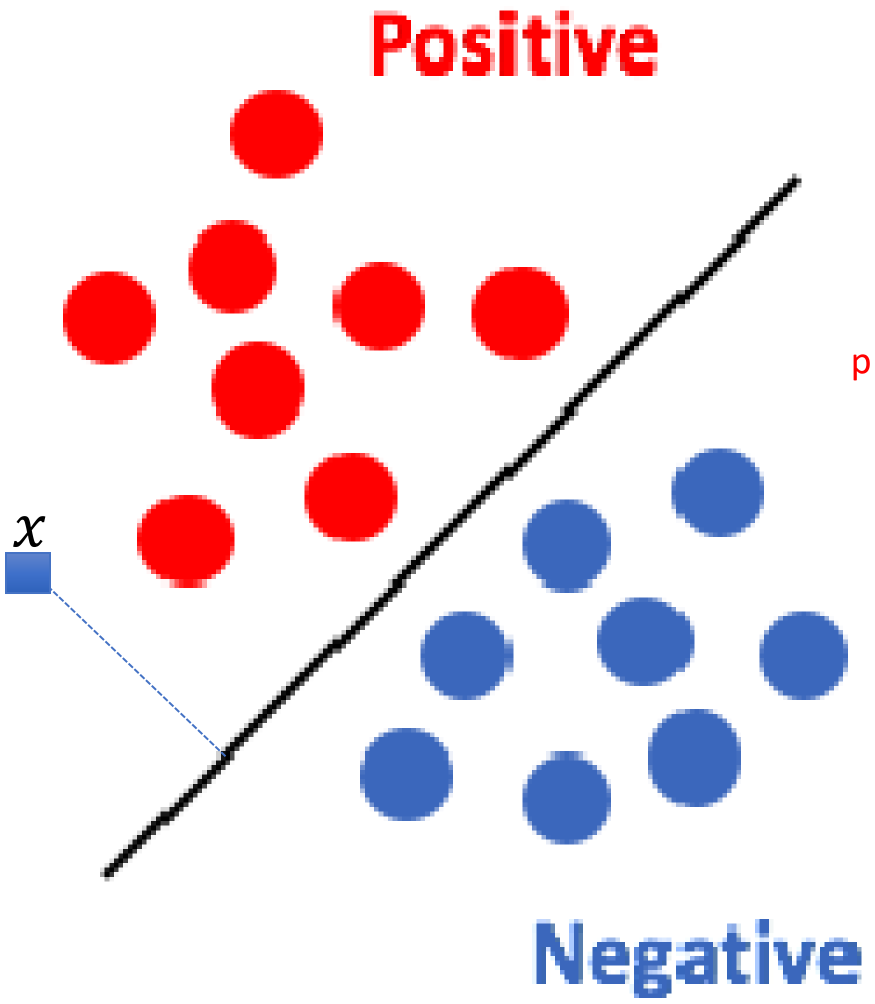
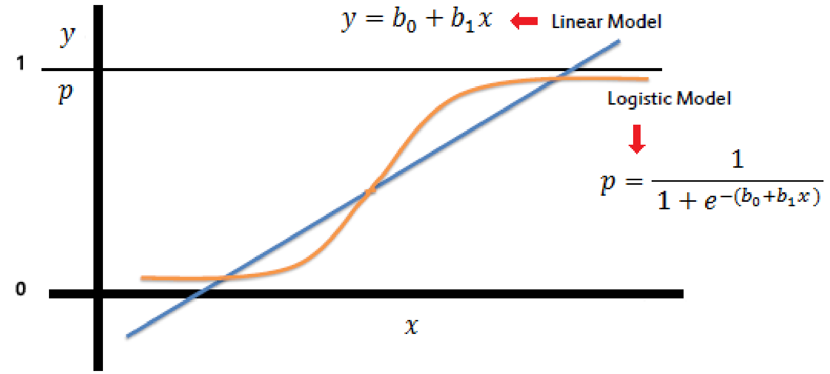
Logistic Regression

Logistic Regression - Intuition





p(X) = p(Y=1|x) = probability that x belongs to positive class

$$\implies p(X) = \frac{e^{(\beta_o + \beta_1 x)}}{e^{(\beta_o + \beta_1 x)} + 1}$$

$$\implies p(e^{(\beta_o + \beta_1 x)} + 1) = e^{(\beta_o + \beta_1 x)}$$

$$\implies p.e^{(\beta_o + \beta_1 x)} + p = e^{(\beta_o + \beta_1 x)}$$

$$\implies p = e^{(\beta_o + \beta_1 x)} - p.e^{(\beta_o + \beta_1 x)}$$

$$\implies p = e^{(\beta_o + \beta_1 x)}(1 - p)$$

$$\implies \frac{p}{1 - p} = e^{(\beta_o + \beta_1 x)}$$

$$\implies \ln(\tfrac{p}{1-p}) = \beta_0 + \beta_1 x$$

Distance of x from decision boundary

Maximum Likelihood

- The likelihood function is the simultaneous density of the observation, as a function of the model parameters.

$$L(\Theta) = \Pr(Data|\Theta)$$

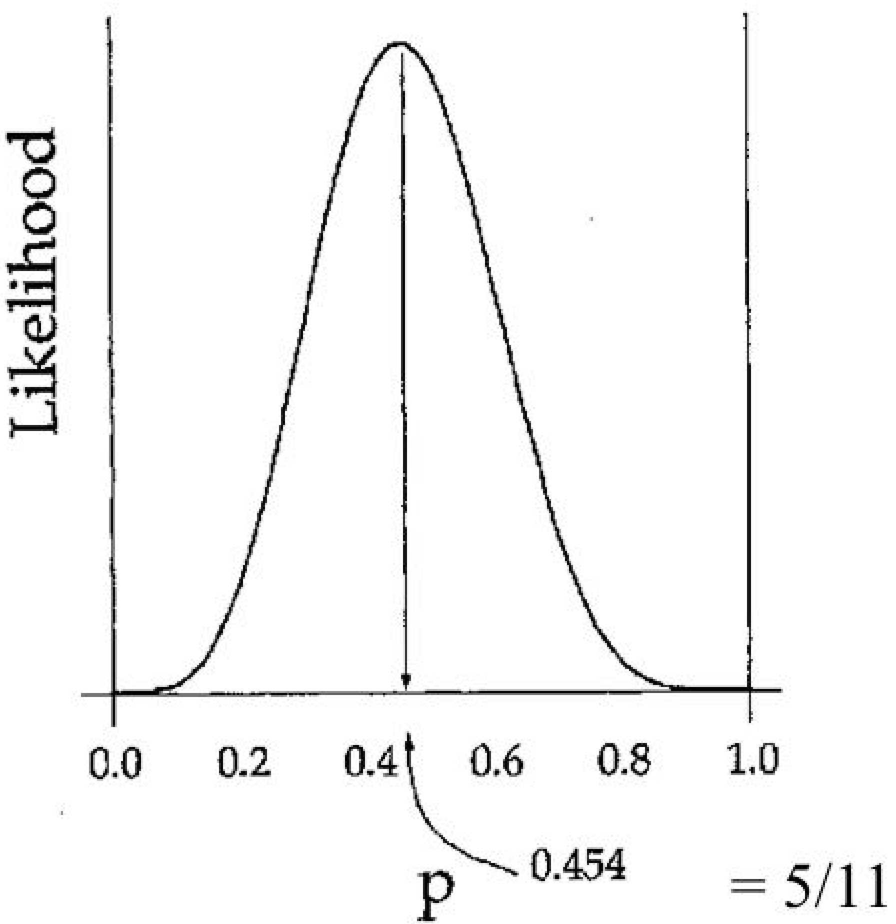
- If the observations are independent, we can decompose the term into

$$\Pr(Data \mid \Theta) = \prod_{i=1}^{n} \Pr(X_i \mid \Theta)$$

An example

- Consider the estimation of heads probability of a coin tossed n times
- Heads probability p
- Data = HHTTHTHHTTT
- $L(p) = \Pr(D|p) = pp(1-p)(1-p)p(1-p)pp(1-p)(1-p)(1-p) = p^5(1-p)^6$

$$L(p) = p^5(1-p)^6$$



Maximum Likelihood

$L(p) = p^5(1-p)^6$

Take the derivative of L with respect to p:

$$\frac{dL}{dp} = 5p^4(1-p)^6 - 6p^5(1-p)^5$$

Equate it to zero and solve:

$$\hat{p} = 5/11$$

Log Likelihood

$L(p) = p^5(1-p)^6$

- For computational reasons, we maximise the logarithm

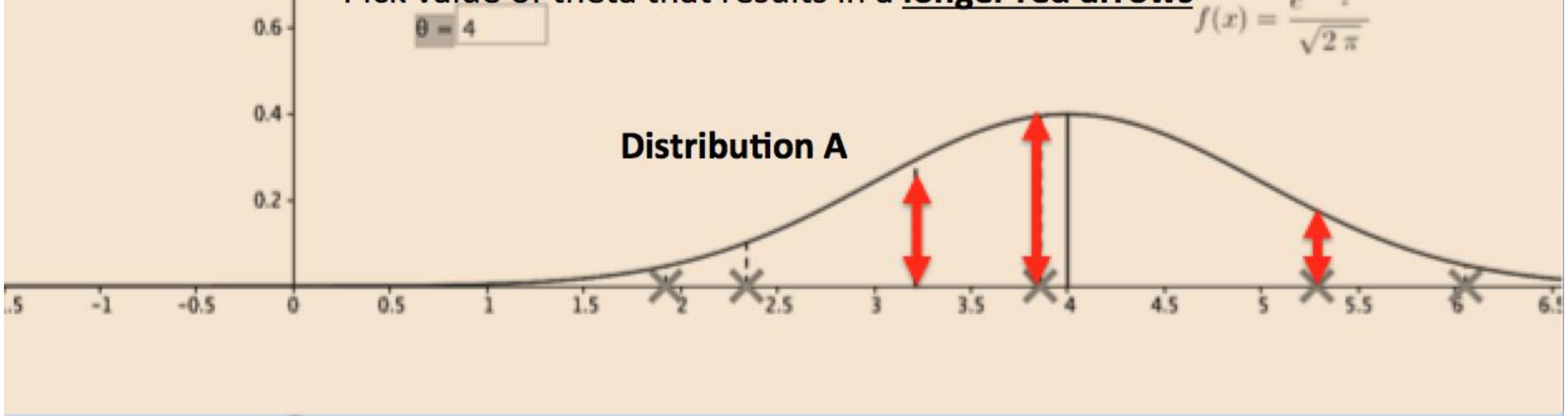
  $\ln L = 5 \ln p + 6 \ln(1-p)$

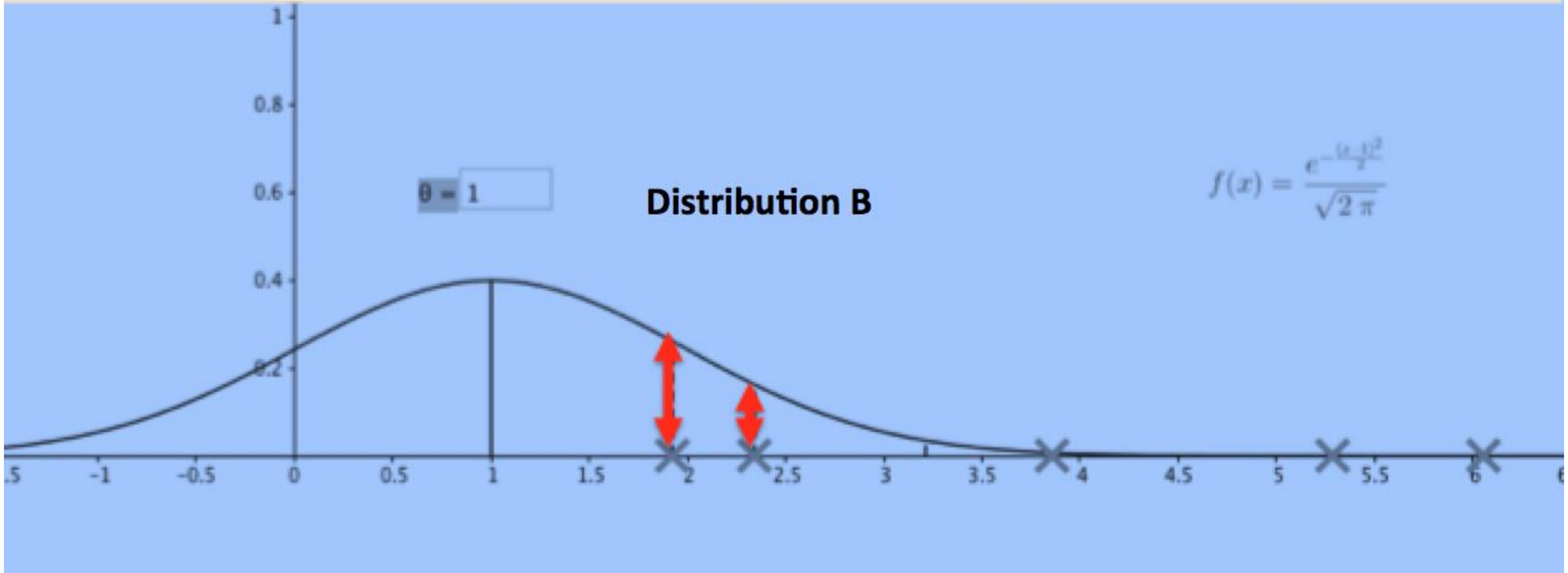
  with derivative

$$\frac{d(\ln L)}{dp} = \frac{5}{p} - \frac{6}{(1-p)} = 0$$

$$\hat{p} = 5/11$$

Maximum Likelihood Estimator will maximize p(x1)p(x2)p(x3)…
-> called the Likelihood function
Pick value of theta that results in a **<u>longer red arrows</u>**





Logistic Regression - Learning parameters

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

$$p(y \mid x; \theta) = (h_\theta(x))^y \,(1 - h_\theta(x))^{1-y}$$

$$\begin{aligned}
L(\theta) &= p(\vec{y} \mid X; \theta) \\
&= \prod_{i=1}^{m} p(y^{(i)} \mid x^{(i)}; \theta) \\
&= \prod_{i=1}^{m} \left(h_\theta(x^{(i)})\right)^{y^{(i)}} \left(1 - h_\theta(x^{(i)})\right)^{1-y^{(i)}}
\end{aligned}$$

$$\begin{aligned}
\ell(\theta) &= \log L(\theta) \\
&= \sum_{i=1}^{m} y^{(i)} \log h(x^{(i)}) + (1 - y^{(i)}) \log(1 - h(x^{(i)}))
\end{aligned}$$

Logistic Regression - Learning parameters

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

$$\begin{aligned}\ell(\theta) &= \log L(\theta) \\ &= \sum_{i=1}^{m} y^{(i)} \log h(x^{(i)}) + (1 - y^{(i)}) \log(1 - h(x^{(i)}))\end{aligned}$$

No closed form.
Cannot set to 0 and solve for θ

$$\frac{\partial}{\partial \theta_j} \ell(\theta) = (y - h_\theta(x))\, x_j$$

Multinomial Logistic Regression

$$\Pr(Y_i = 1) = \frac{e^{\boldsymbol\beta'_1 \cdot \mathbf{X}_i}}{1 + \sum_{k=1}^{K-1} e^{\boldsymbol\beta'_k \cdot \mathbf{X}_i}}$$

$$\cdots\cdots$$

$$\Pr(Y_i = K-1) = \frac{e^{\boldsymbol\beta'_{K-1} \cdot \mathbf{X}_i}}{1 + \sum_{k=1}^{K-1} e^{\boldsymbol\beta'_k \cdot \mathbf{X}_i}}$$

$$\Pr(Y_i = K) = \frac{1}{1 + \sum_{k=1}^{K-1} e^{\boldsymbol\beta'_k \cdot \mathbf{X}_i}}$$