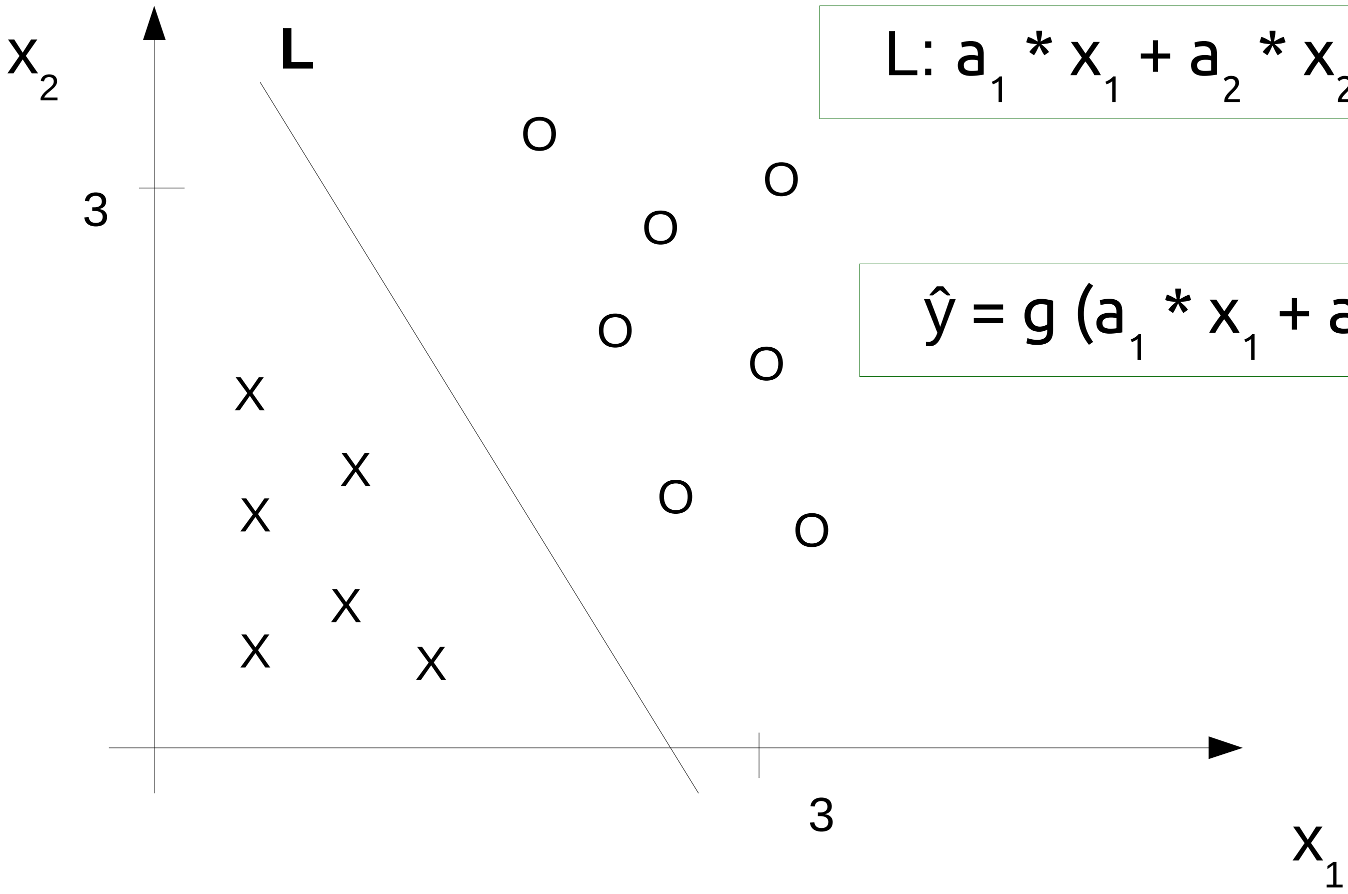
Logistic Regression

Example - Two Features



$$L: a_1 * x_1 + a_2 * x_2 + b = 0$$

$$\hat{y} = g\,(a_1 * x_1 + a_2 * x_2 + b)$$

Explanation

- L is not the line of best fit as in linear regression.
- It is a line dividing regions where $a_1 * x_1 + a_2 * x_2 + b$ is < and > 0.
- It is the line of best separation.
- Lets call $f(x) = a_1 * x_1 + a_2 * x_2 + b$
- yhat = g(f(x)

Explanation

- the further to the right of the line you are
  - f(x) gets larger
  - g(f(x)) gets closer to 1
- The further to the left of the line you are
  - f(x) gets larger negative
  - g(f(x)) gets closer to 0

Solution for p Dependent Variables

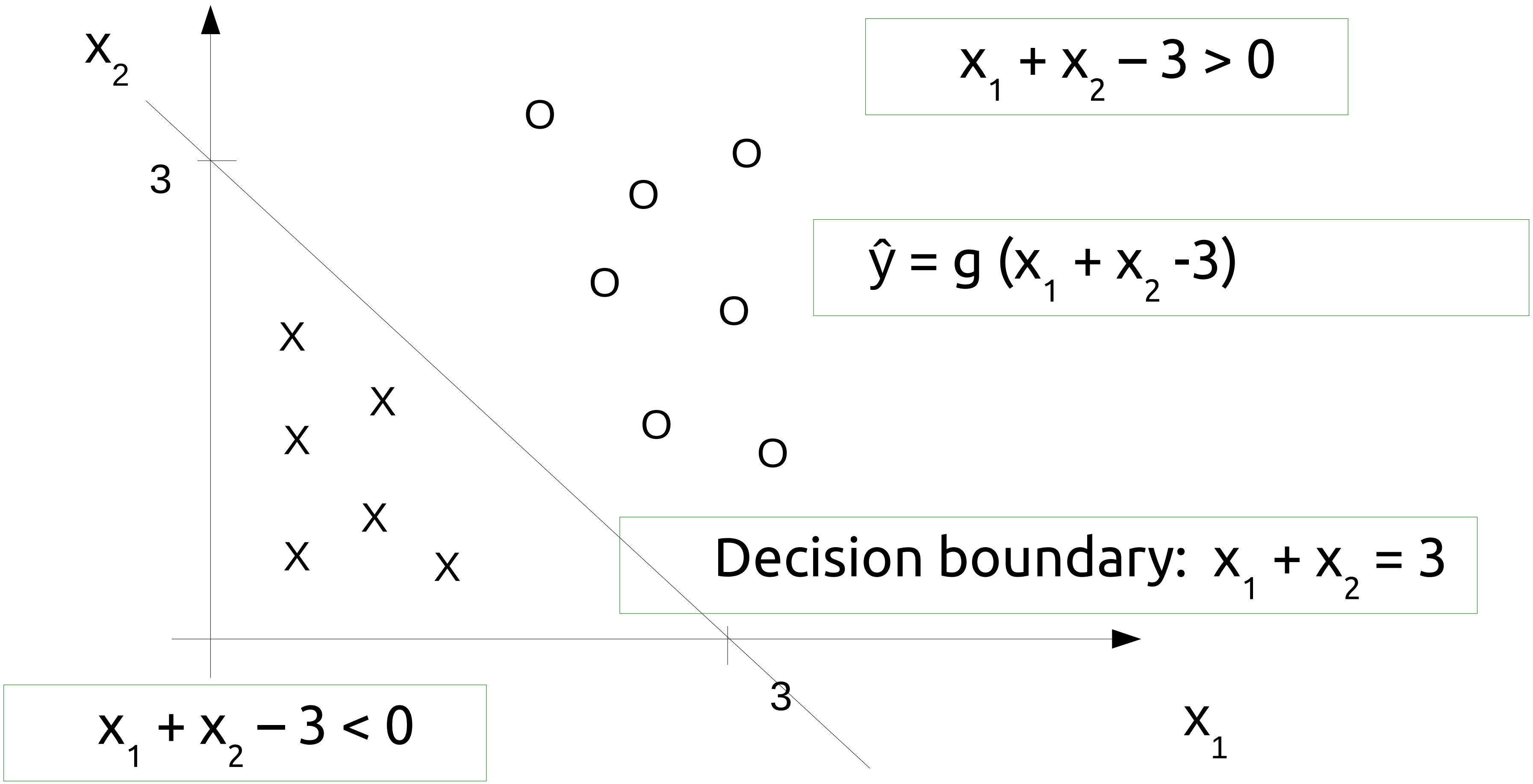
$$\boxed{Estimate\ of\ P\left(y=1 \mid x_1 .. x_n\right)=g\left(f\left(x\right)\right)}$$

$$\boxed{Estimate\ of\ P\left(y=1 \mid x_1 .. x_n\right)=\frac{1}{\left(1+e^{-\left(\sum_k a_k x_k+b\right)}\right)}}$$

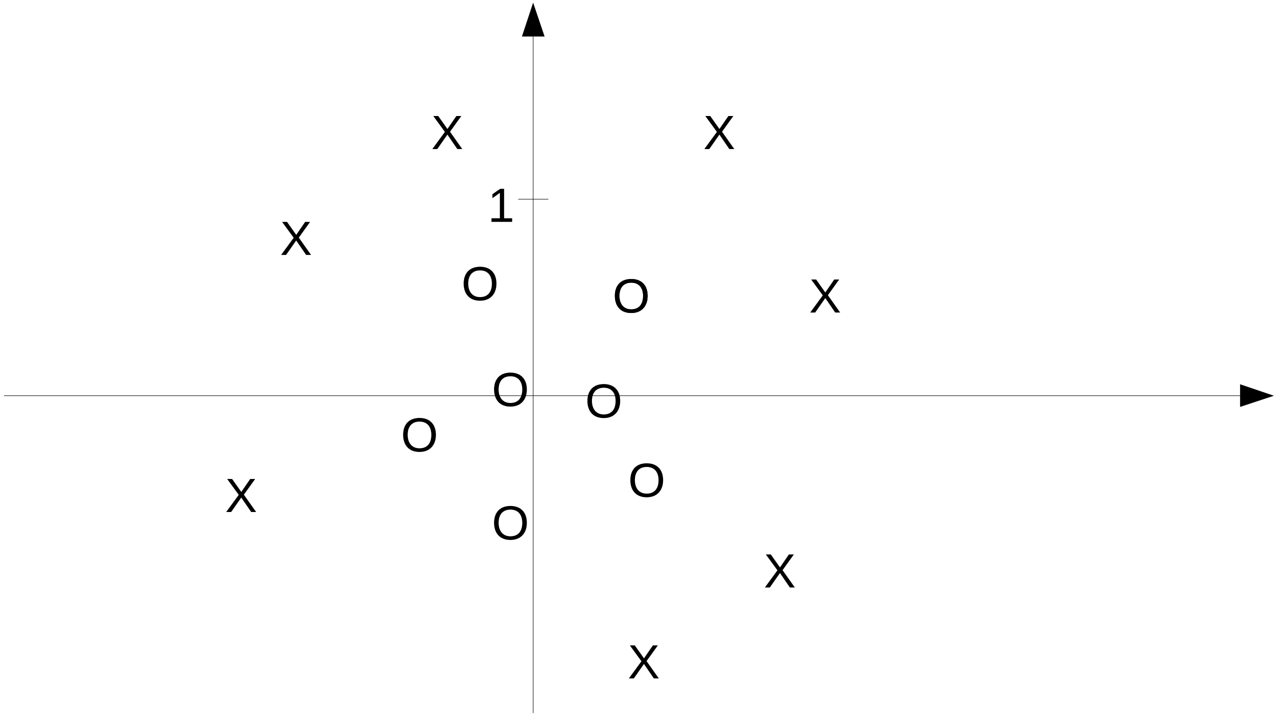
Example - Decision Boundary

- [Finding a_1 a_2 and b later.]
- Suppose $a_1=1$ and $a_2=1$ and $b = -3$
- Predict 1 when
  - $x_1 + x_2 - 3 >= 0$
  - $x_1 + x_2 >= 3$

Example - Decision Boundary



Non Linear Decision Boundaries



Non Linear Decision Boundaries

- Add non linear terms to the regression equation.
- $\hat{y} = g\,(a_1{*}x_1 + a_2{*}x_2 + a_3{*}x_1^2 + a_4 * x_2^2 + b)$
- Get
  - $a_1 = 0$, $a_2=0$, $a_3*= 1$, $a_4=1$, $b = -1$
- Decision boundary
  - $x_1^2 + x_2^2 \geq 1$
- Add more complex terms (e.g. $x_1^2 * x_2^2$)to get more complicated decision boundaries.

Non Linear Decision Boundaries