# Logistic Regression

## Finding the Parameters

# Finding the Parameters

- Like linear regression, this is done by optimizing a cost function.
- By minimizing a cost function.
- Cost function sometimes called an objective function.

# Finding the Parameters

- n - number of factors
- m - number of training instances
- $y^i$ - y values for the training set
- $\hat{y}^i$ - predicted y values for the training set
- parameters a, $b_1$, $b_2$, ..

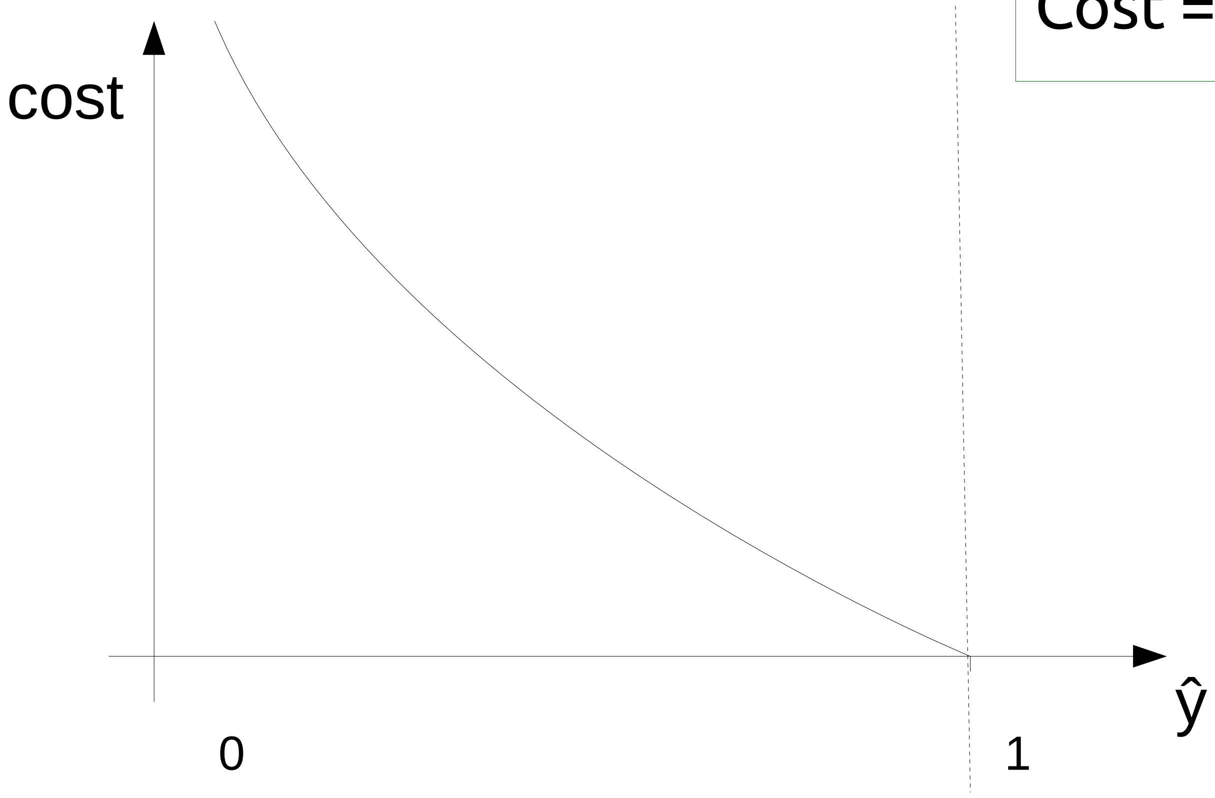
# Cost Function for Logistic Regression

$$
\begin{gathered}
cost(\hat{y}, y) = -\log \hat{y} \; if \; y = 1 \\
cost(\hat{y}, y) = -\log(1 - \hat{y}) \; if \; y = 0
\end{gathered}
$$

- Recall ŷ is between 0 and 1

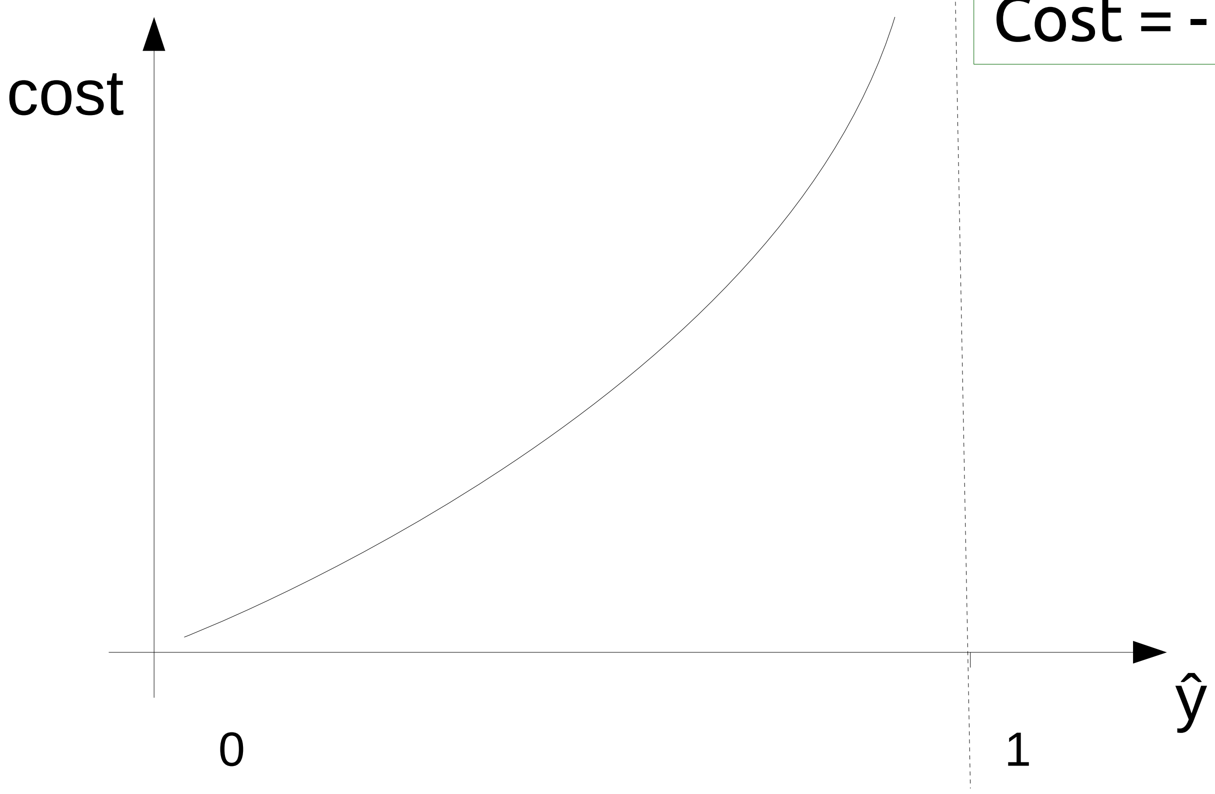
# Cost function

Cost = - log ŷ  if y = 1



# Cost function

$$\text{Cost} = -\log(1 - \hat{y}) \quad \text{if } y = 0$$



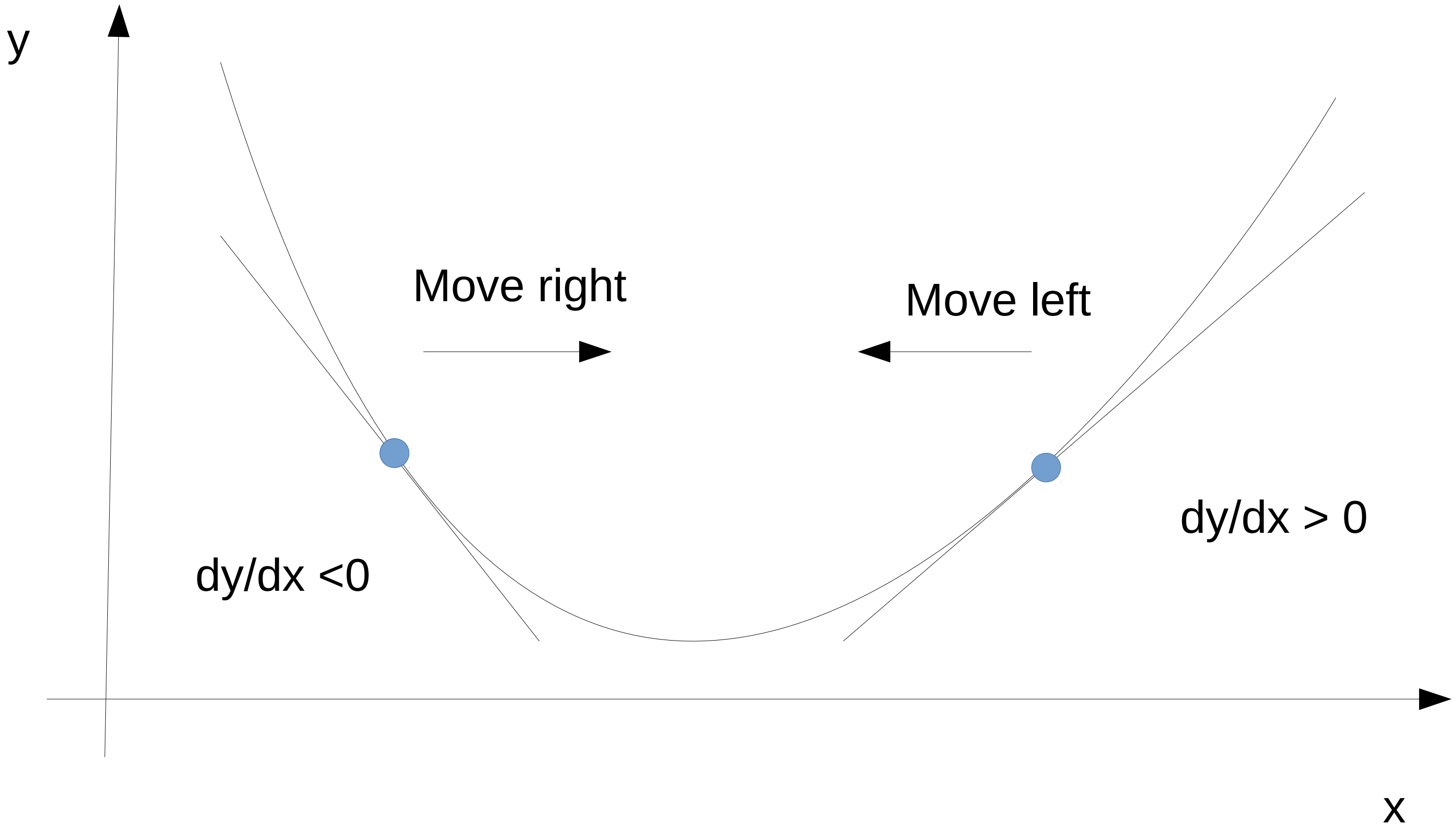
# Cost function

- If y = 1 and we predict a value close to 0, then there is a heavy cost (penalty)
- If y = 0 and we predict a value close to 1, then there is also a heavy cost (penalty)

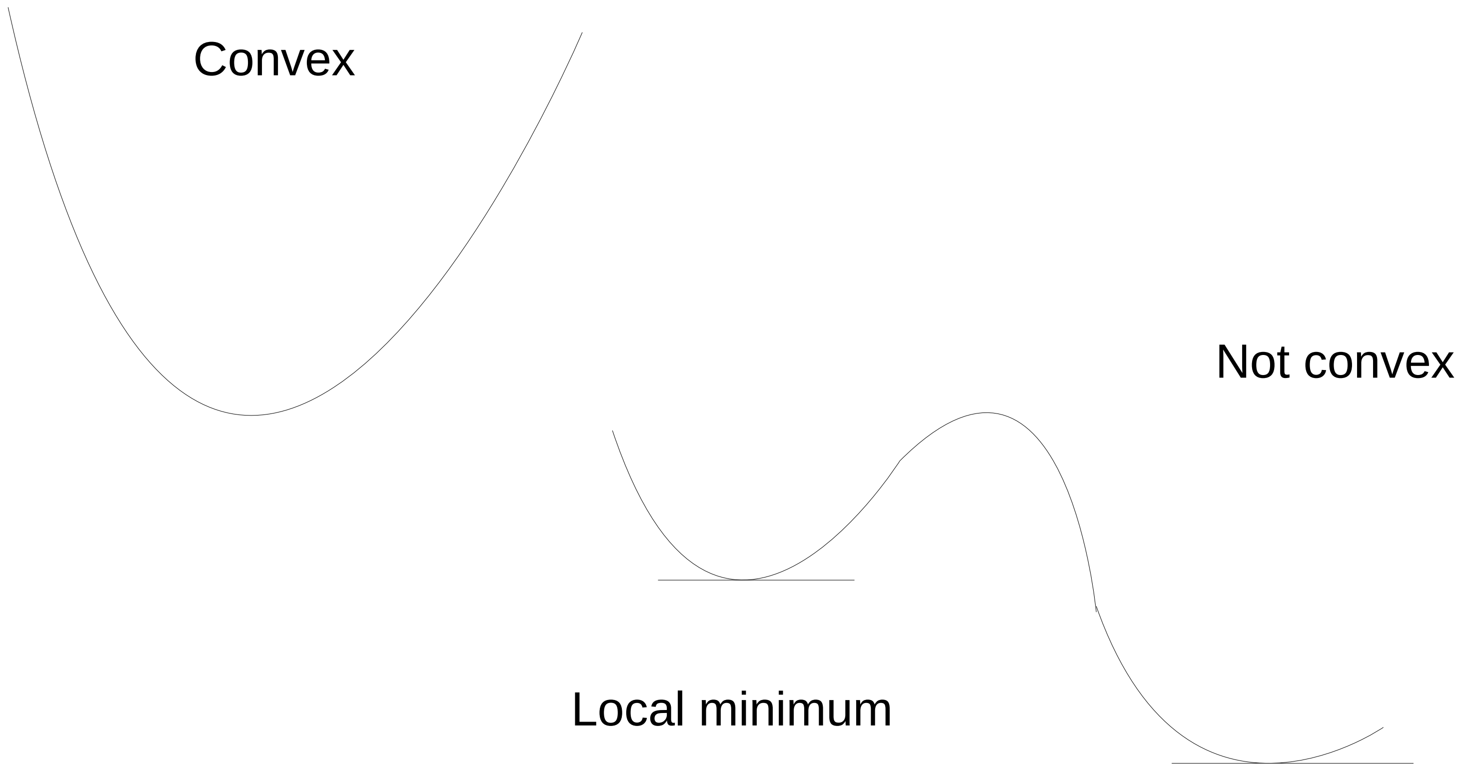
# Gradient Descent

- Gradiant Descent
- If dy/dx is negative move right
- If positive move left.
- Finds a local minimum.
- It turns out that the cost function defined above is convex, so gradient descent finds a absolute minimum.

# Gradient Descent



# Convex Cost Function



# Gradient Descent

- Gradient Descent is used to find the values of the parameters a and b's that minimize the value of the cost function for the training data.
- This is implemeted  in R .

# Logistic Regression - Summary

- Find the equation of the line that best seperates the classes.
  - $z = a_1 * x_1 + a_2 * x_2 \ldots\ldots\ldots\ldots + a_n * x_n + b$
  - This is done by minimising the cost function defined above.
- To predict the class of a new instance:-
  - Calculate the values of z given above
  - Find g(z) where g is the logistic function.
  - This gives the probability (of a positive).