Master in Internet of Things for eHealth

M5. Smart Data Knowledge / Analytics

Logistic Regression

Instructor **David Gerónimo**
research@davidgeronimo.com

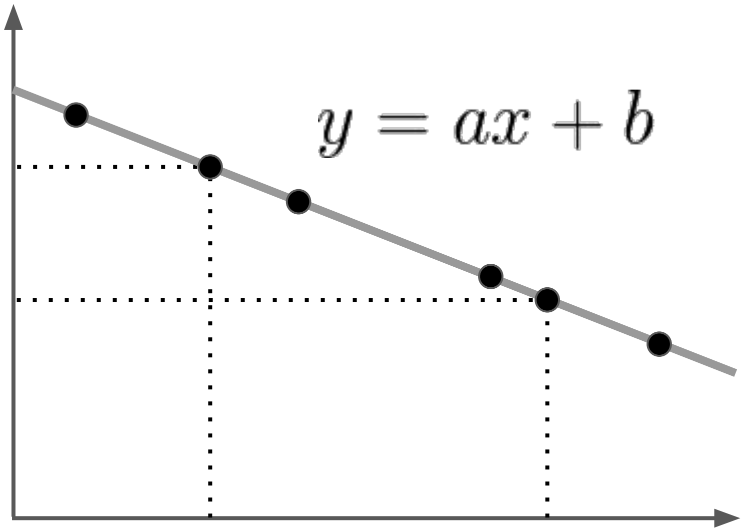
Universitat Autònoma de Barcelona

January 8th 2019

Linear Regression

- We want to estimate a line defined by *a* and *b* that relates *x* and *y*.

Car Insurance Price



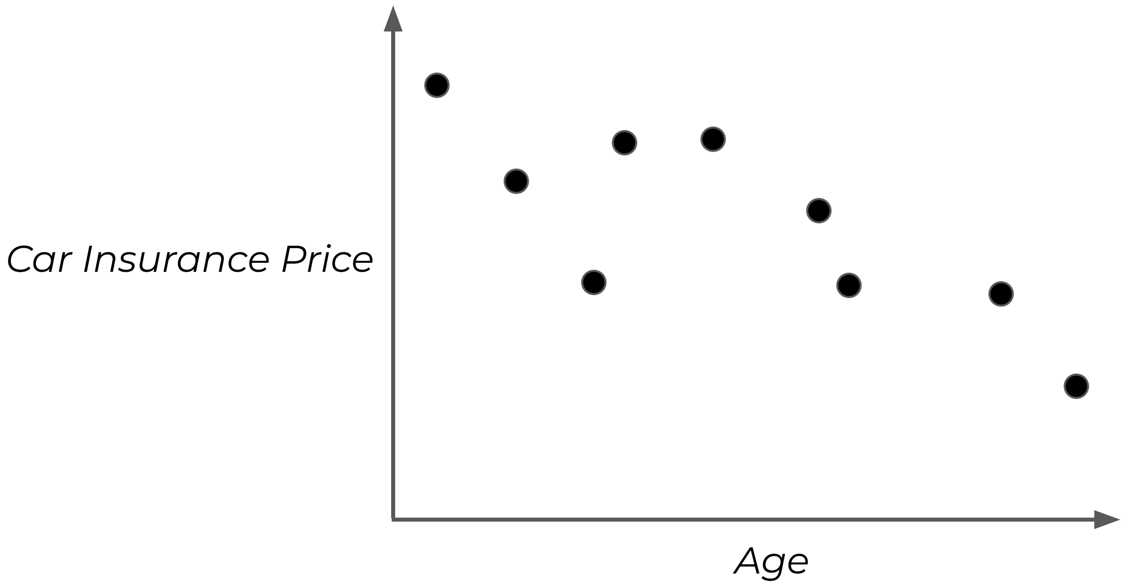
Age

Parameters $\theta = \begin{bmatrix} a \\ b \end{bmatrix}$

Perfect linear relationship

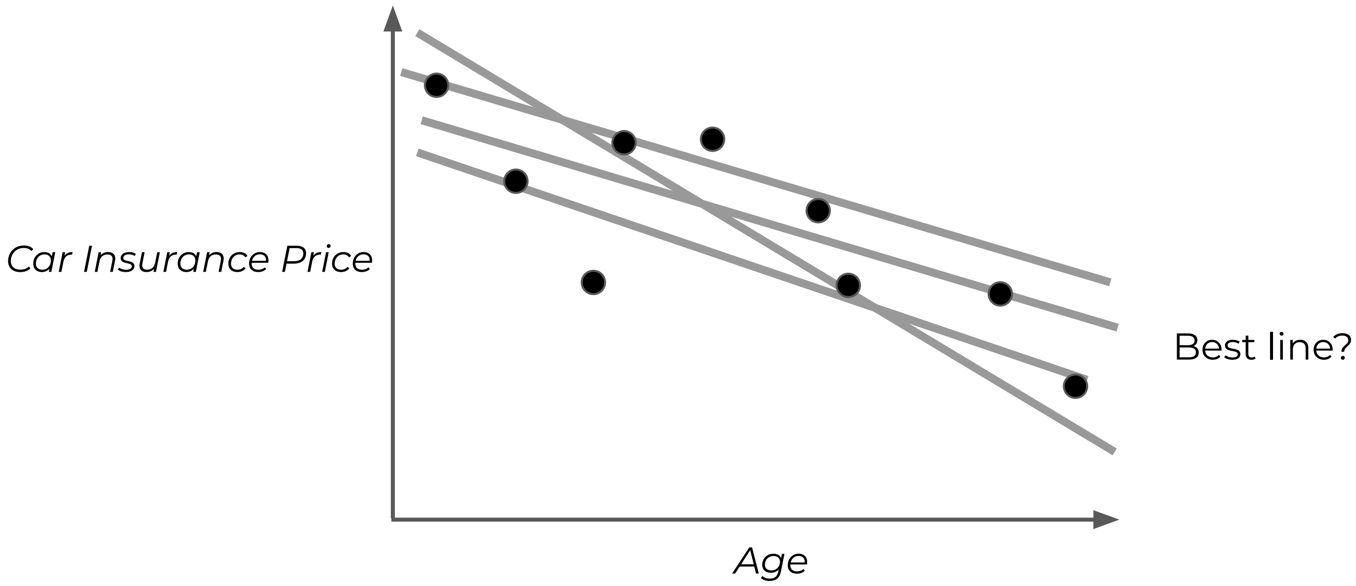
Linear Regression

- In real life, such a perfect linear relationship rarely happens



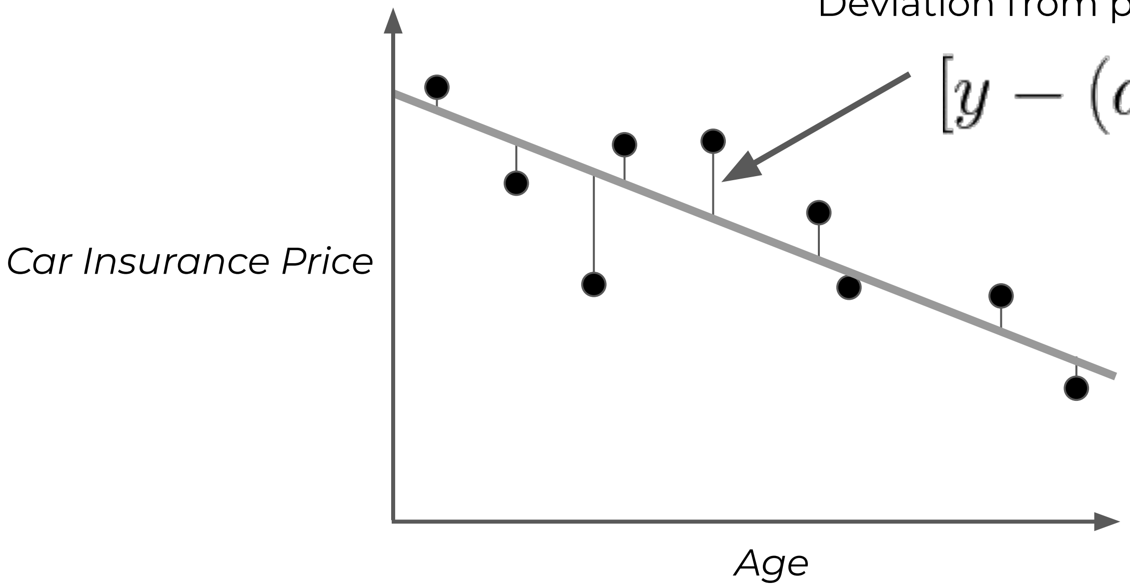
Linear Regression

- In real life, such a perfect linear relationship rarely happens



Linear Regression

- Linear Least Squares

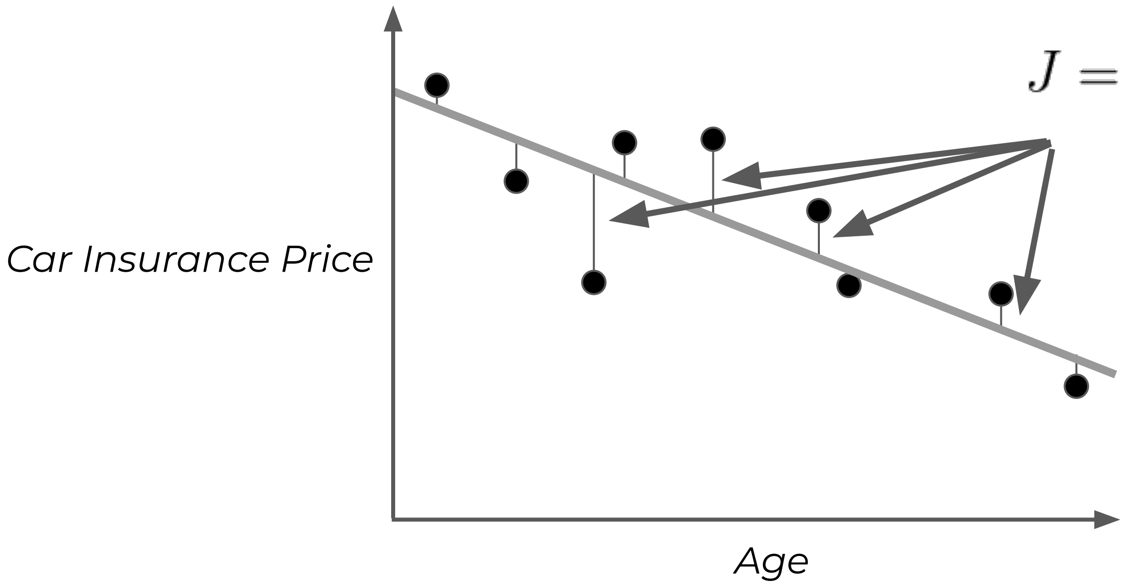


Deviation from point to line:

$$[y - (ax + b)]^2$$

Linear Regression

- Linear Least Squares



We want to minimize the following **cost function** (called **Mean Squared Error**):

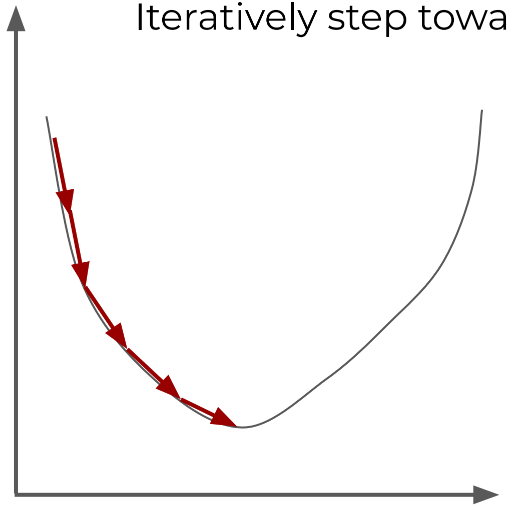
$$J = \sum_{i=1}^{n} [y_i - (ax_i + b)]^2$$

Linear Regression

- How do we **optimize** (i.e., find the minimum J)? ⇒ by **Gradient Descent**

Iteratively step towards the derivative of J

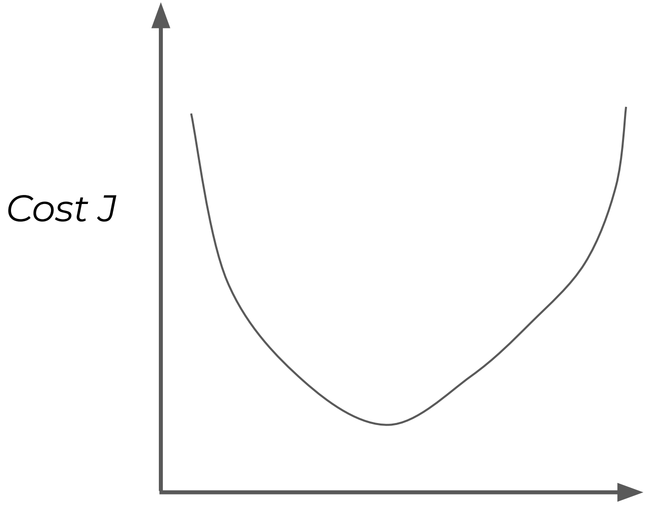
Cost J



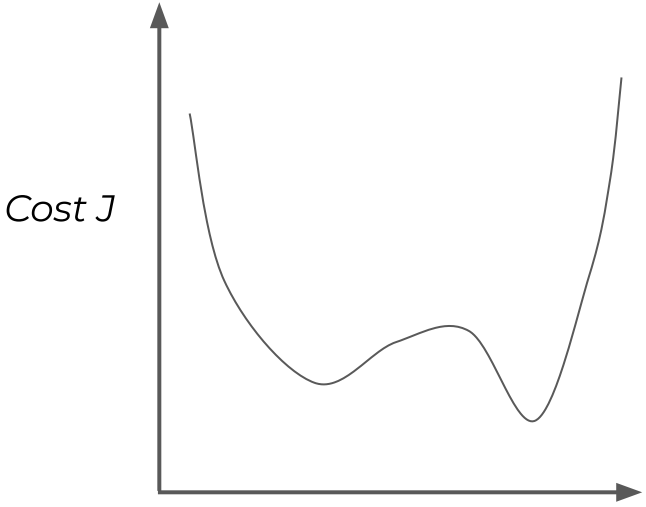
Convex

Linear Regression

- We can use gradient descent because J is convex



Convex



Non-Convex

Linear Regression

- Derivative of J (partial derivatives wrt *a* and *b*)

$$\frac{\partial J}{\partial a} = \frac{2}{n}\sum_{i=1}^{n}(y_i - (ax_i + b))x_i$$

$$\frac{\partial J}{\partial b} = \frac{2}{n}\sum_{i=1}^{n}(y_i - (ax_i + b))$$

Linear Regression

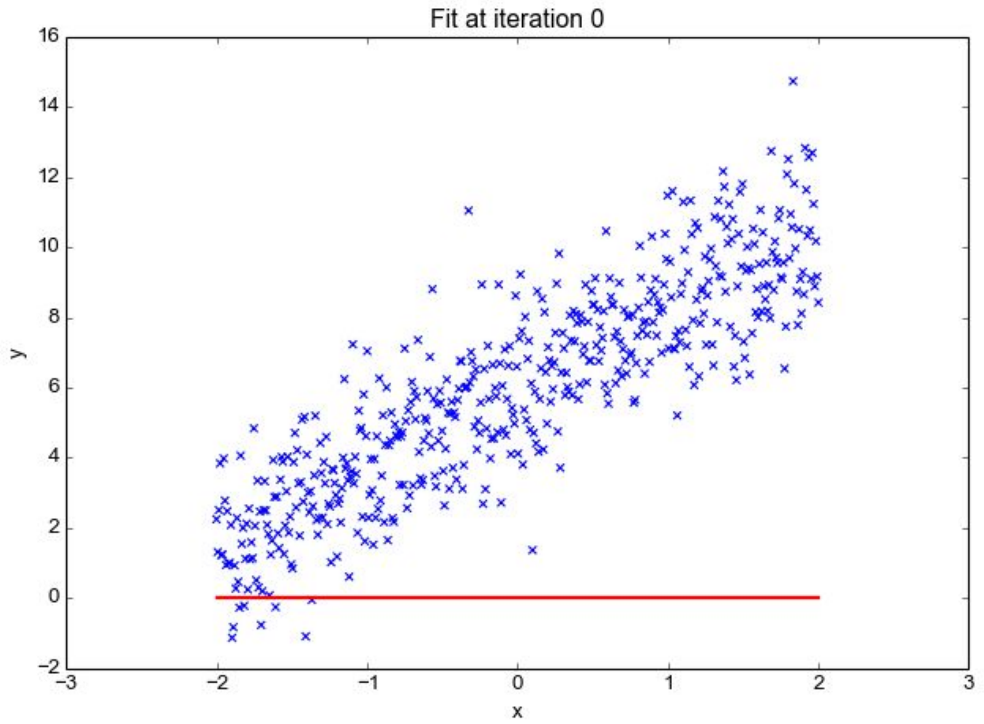
- Descent iteratively through the cost using the gradient at a rate α

$$a = a - \alpha \frac{2}{n} \sum_{i=1}^{n} (y_i - (ax_i + b))x_i$$

$$b = b - \alpha \frac{2}{n} \sum_{i=1}^{n} (y_i - (ax_i + b))$$

Linear Regression

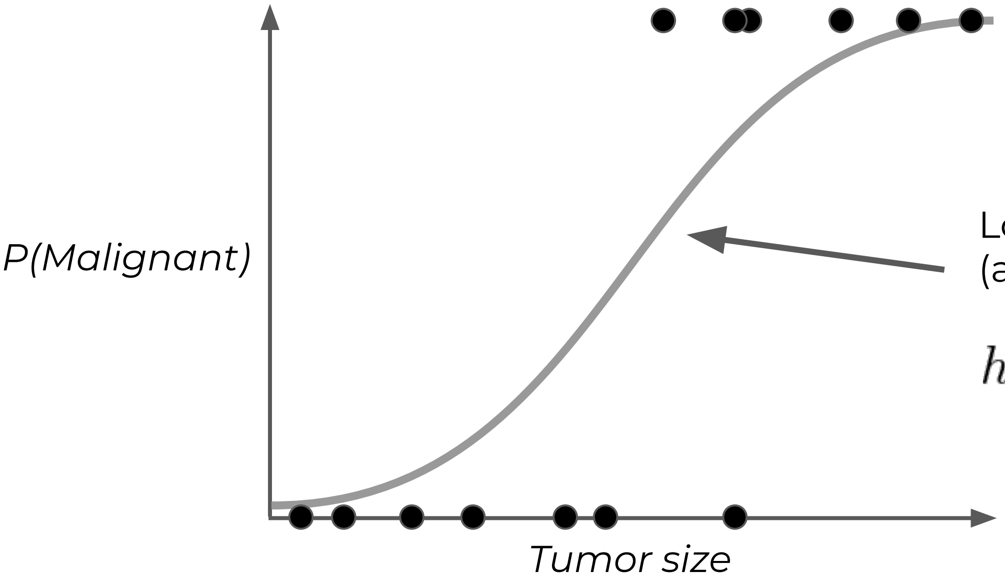
- Optimization result



Source: towardsdatascience.com

Logistic Regression

- We want to estimate *x* and *y* but now *y* is categorical (1 or 0) instead of linear.

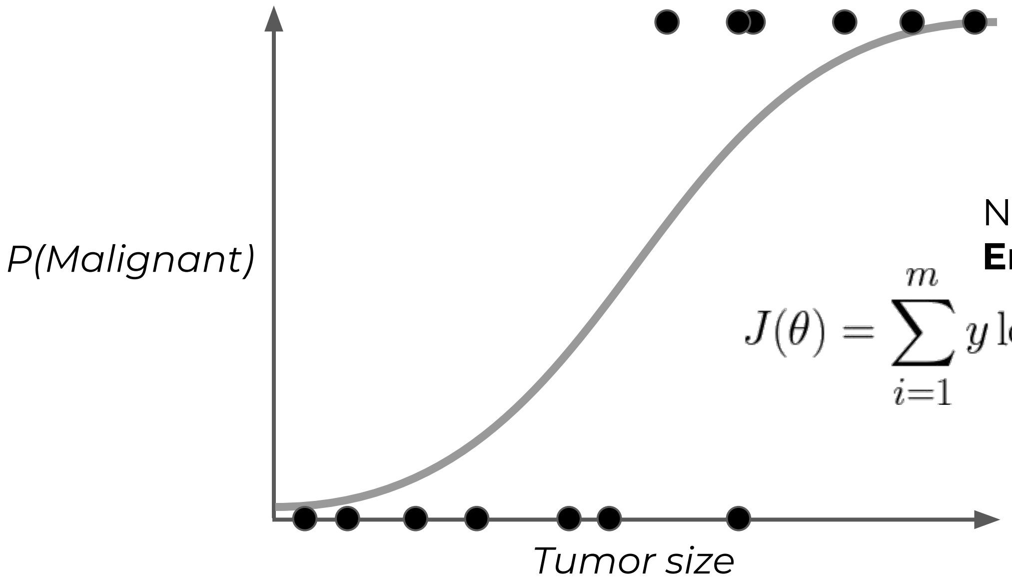


Logistic function
(a.k.a. Sigmoid)

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

Logistic Regression

- We want to estimate *x* and *y* but now *y* is categorical (1 or 0) instead of linear.



Now the **cost function** (**Cross Entropy**) is more esoteric :

$$J(\theta) = \sum_{i=1}^{m} y \log(h_\theta(x)) - (1-y)\log(1-h_\theta(x))$$

Logs added to make the problem convex

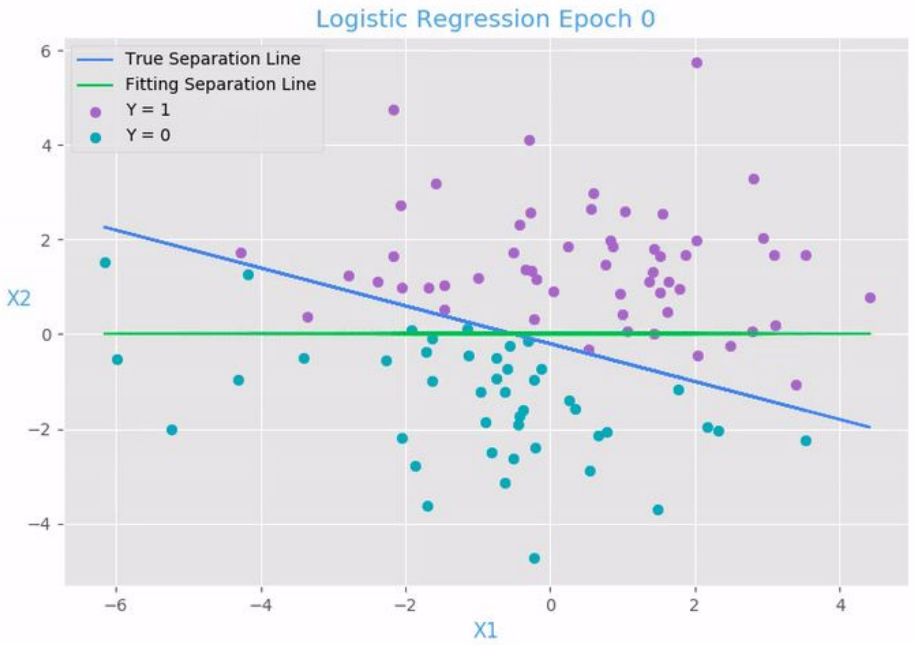
Logistic Regression

- Derivative of the cost function (using the chain rule, the derivative of the sigmoid and the constant $\frac{\partial z}{\partial \theta} = x$ where $z = \theta^T x$)

$$\frac{\partial}{\partial \theta_j} J(\theta) = \sum_{i=1}^{m} (h_\theta(x^i) - y^i) x_j^i$$

Logistic Regression

- Optimization result



Source: towardsdatascience.com