Classification: Logistic Regression

Dr. Chandranath Adak

Dept. of CSE, Indian Institute of Technology Patna

September 3, 2025

Introduction

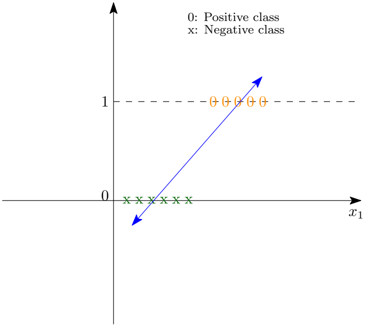
1. Supervised learning
2. It is a classification technique having discrete class labels
3. Objective: To classify a sample represented by a set of features
4. In classification, the aim is to find the class label given a set of input features $X \in \mathbb{R}^k$

Examples

- Given a set of emails, classify them as spam / not-spam
- Given a set of images of animals, classify them as dog / cat
- Given barometer's reading, predict the weather as rainy / not rainy

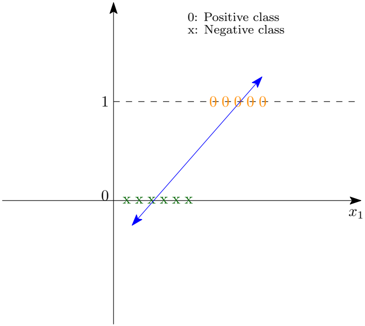
How to extend linear regression for classification?

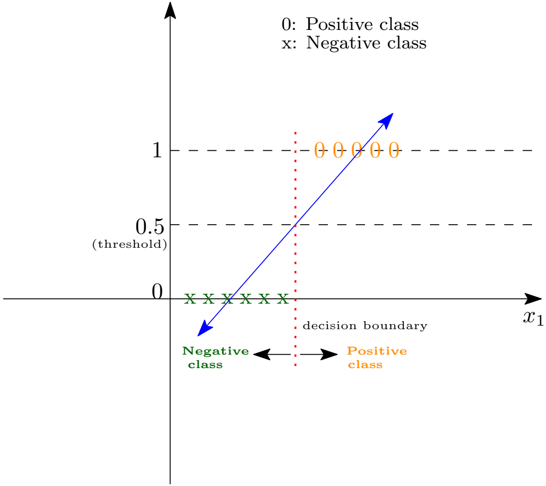
- Apply a threshold on regression output h_θ



How to extend linear regression for classification?

- Apply a threshold on regression output h_θ



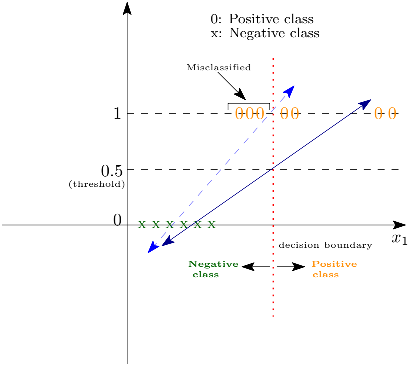


Is the classical linear regression formulation is sufficient for classification?

- The span of h_θ is from $-\infty$ to ∞
- Therefore, the value of h_θ can be much greater than 1 or less than 0

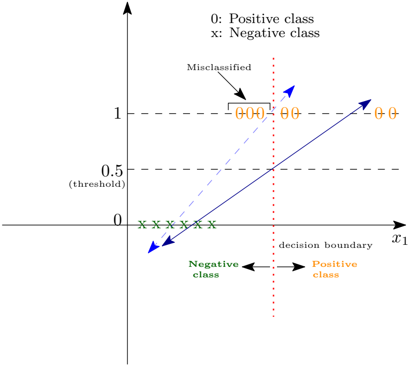
Is the classical linear regression formulation is sufficient for classification?

- The span of h_θ is from $-\infty$ to ∞
- Therefore, the value of h_θ can be much greater than 1 or less than 0



Is the classical linear regression formulation is sufficient for classification?

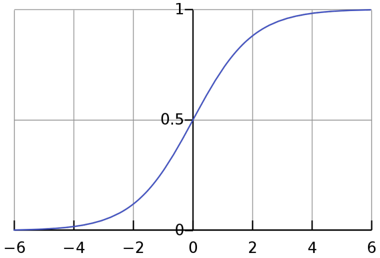
- The span of h_θ is from $-\infty$ to ∞
- Therefore, the value of h_θ can be much greater than 1 or less than 0



Choosing a threshold is challenging

Can we bound the hypothesis value?

- Use Sigmoid function (also termed as logistic function) $\sigma(z) = \frac{1}{1+e^{-z}}$



- Characteristics of the sigmoid function:
  - $0 \leq \sigma(z) = \frac{1}{1+e^{-z}} \leq 1$
  - $\sigma(0) = 0.5$
- The input to the sigmoid is $z = \sum_{i=0}^{k} \theta_i x_i$
  (for k features, considering linear function and $x_0 = 1$)

Class boundaries

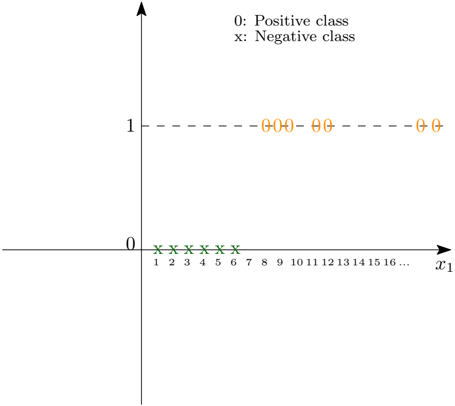
- $\sigma(z) = 0.5$ can be used as the threshold value

$$\text{Sample } X(x_1, x_2, \ldots, x_k) \in \begin{cases} \text{Positive class} & \text{, if } \sigma(z) \geq 0.5 \\ \text{Negative class} & \text{, otherwise} \end{cases}$$

which implies

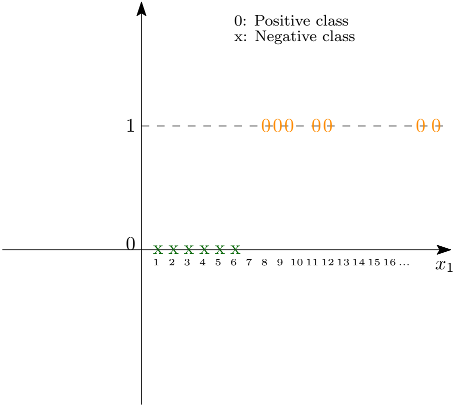
$$\text{Sample } X(x_1, x_2, \ldots, x_k) \in \begin{cases} \text{Positive class} & \text{, if } z \geq 0 \\ \text{Negative class} & \text{, otherwise} \end{cases}$$

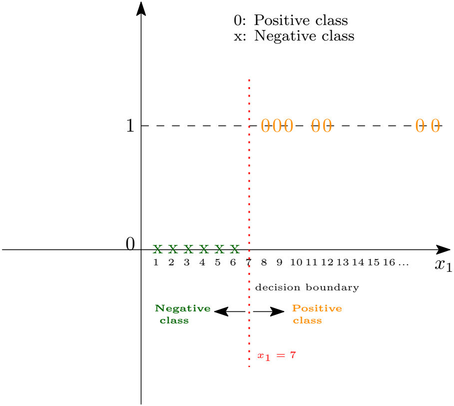
Class boundaries: examples



- $z = \theta_0 + \theta_1 x_1$, we need to predict θ_0 and θ_1
- Assume $\theta_0 = -7$ and $\theta_1 = 1$

Class boundaries: examples

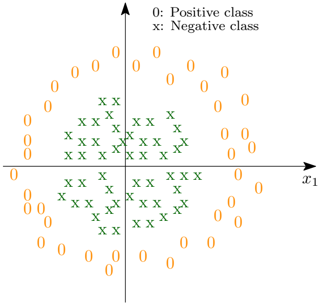




- $z = \theta_0 + \theta_1 x_1$, we need to predict θ_0 and θ_1
- Assume $\theta_0 = -7$ and $\theta_1 = 1$

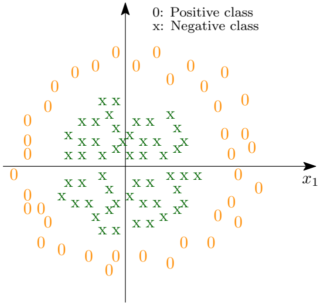
- Then $z = -7 + x_1$
- For $\sigma(z) \geq 0.5 \rightarrow z \geq 0 \rightarrow x_1 \geq 7$
- Class boundary $x_1 = 7$

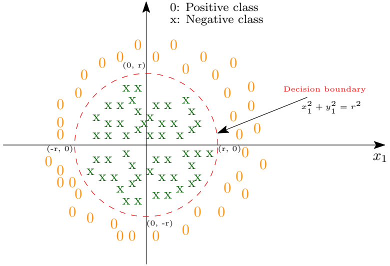
Class boundaries: examples



- $z = \theta_0 + \theta_1 x_1^2 + \theta_2 x_2^2 + \theta_3 x_1 x_2 + \theta_4 x_1 + \theta_5 x_2$, we need to predict $\theta_0, \theta_1, \ldots, \theta_5$
- Assume $\theta_0 = -r^2$ and $\theta_1 = 1, \theta_2 = 1$ and rest are 0

Class boundaries: examples





- $z = \theta_0 + \theta_1 x_1^2 + \theta_2 x_2^2 + \theta_3 x_1 x_2 + \theta_4 x_1 + \theta_5 x_2$, we need to predict $\theta_0, \theta_1, \ldots, \theta_5$
- Assume $\theta_0 = -r^2$ and $\theta_1 = 1, \theta_2 = 1$ and rest are 0
- Then $z = -r^2 + x_1^2 + x_2^2$
- For $\sigma(z) \geq 0.5 \rightarrow z \geq 0 \rightarrow x_1^2 + x_2^2 \geq r^2$
- Class boundary: $x_1^2 + x_2^2 = r^2$

Cost function

- Hypothesis: $h_\theta^i = h_\theta(X^i) = \frac{1}{1+e^{-(\sum_{j=0}^{k} \theta_j x_j^i)}}$
- As per linear regression, the cost function $J_\theta = \frac{1}{2n}\sum_{i=1}^{n}(h_\theta^i - y^i)^2$
- J_θ is no longer a convex function
  - Consisting of multiple local minimas

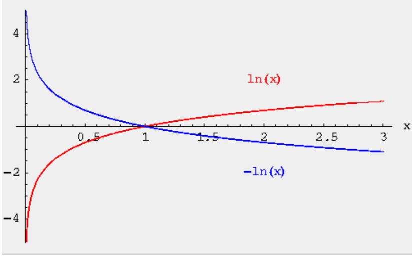
Need a cost function with a single minima

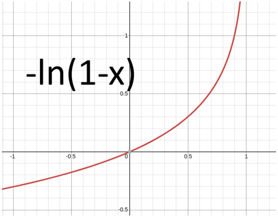
Notice: $J_\theta = \frac{1}{2n}\sum_{i=1}^{n}(h_\theta^i - y_j^i)^2 = \frac{1}{n}\sum_{i=1}^{n} \underbrace{loss(h_\theta^i, y^i)}_{\substack{\text{function of actual value}\\ \text{and predicted value}}}$

Redefining the cost function

$$J_\theta = \frac{1}{2n}\sum_{i=1}^{n}(h_\theta^i - y_j^i)^2 = \frac{1}{n}\sum_{i=1}^{n} \underbrace{loss(h_\theta^i, y^i)}_{\text{function of actual value and predicted value}}$$

$$loss(h_\theta^i, y^i) = \begin{cases} -log(h_\theta^i) & \text{, if } y^i = 1 \\ -log(1 - h_\theta^i) & \text{, if } y^i = 0 \end{cases}$$





Redefining the cost function (contd.)

Combining both the cases:

$$loss(h_\theta^i, y^i) = y^i(-log(h_\theta^i)) + (1 - y^i)(-log(1 - h_\theta^i))$$

Redefining the cost function (contd.)

Combining both the cases:

$$loss(h_\theta^i, y^i) = y^i(-log(h_\theta^i)) + (1 - y^i)(-log(1 - h_\theta^i))$$

Notice:

- When $y^i = 1$: $loss(h_\theta^i, y^i) = -log(h_\theta^i)$
- When $y^i = 0$: $loss(h_\theta^i, y^i) = -log(1 - h_\theta^i)$

If you perform the gradient descent, the updation formula of θ remains same

Try to find the derivative of J_θ with respect to θ_j!

Multi class classification

- If you have multiple classes $C_1, C_2, \ldots, C_m$
- For each class C_i for $i = 1, 2, \ldots, m$, classify as C_i versus rest
- Classify a sample X in class $\max_{j=1,2,\ldots,m}(h_\theta^j(X))$

Thank You!