Assumptions of Logistic Regression

05 July 2023 07:21

log odd → logit

Y → bernoulli dist → MLE → likelihood

Logistic regression, like other statistical methods, relies on certain assumptions. Here are the main assumptions of logistic regression:

1. Binary Logistic Regression requires the dependent variable to be binary: That means the outcome variable must have two possible outcomes, such as "yes" vs "no", "success" vs "failure", "spam" vs "not spam", etc.
2. Independence of observations: The observations should be independent of each other. In other words, the outcome of one instance should not affect the outcome of another.
3. Linearity of independent variables and log odds: Although logistic regression does not require the dependent and independent variables to be related linearly, it requires that the independent variables are linearly related to the log odds.
4. Absence of multicollinearity: The independent variables should not be too highly correlated with each other, a condition known as multicollinearity. While not a strict assumption, multicollinearity can be a problem because it can make the model unstable and difficult to interpret.
5. Large sample size: Logistic regression requires a large sample size. A general guideline is that you need at least 10 cases with the least frequent outcome for each independent variable. For example, if you have 5 independent variables and the expected probability of your least frequent outcome is 0.10, then you would need a minimum sample size of 500 (10*5 / 0.10).

Note that violating these assumptions doesn't mean you can't or shouldn't use logistic regression, but it may impact the validity of the results and you should proceed with caution.

Y: 1, 0, 1, 0

MLE

→ linear → $X_1 X_2 X_3 \ldots$

→ large sample size → inference

$Y \to \beta_1$

$\boxed{\beta_1\ \beta_2 - \beta_n}$

$\boxed{10\times5}$

$\frac{5\times10}{0.10} \to \boxed{500} \to$ rows

$x_1\ x_2\ x_3\ x_4\ x_5 \mid y$

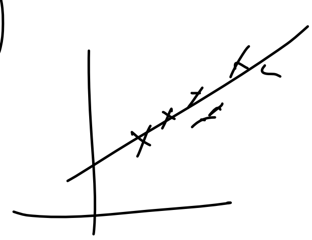
1 → 90%

0 → 10%

(0) → 10% . 0 90%

linear reg →

X | Y



$\boxed{Y = \beta X} \leftarrow Y = \beta X$

linear

X | Y

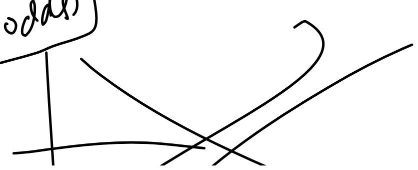
non linear

$\boxed{Y = \frac{1}{1+e^{-\beta X}}}$

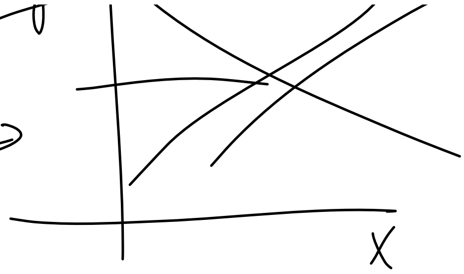
X → log(odds)

linear

$\boxed{\log(\text{odds})}$



linear



$\boxed{\text{Odds}} \rightarrow$ log(odds)

$$\log\left(\frac{p}{1-p}\right) = \beta X$$

Odds and Log(Odds)

05 July 2023 07:20

prob →

Odds: The odds of an event is the ratio of the probability of the event happening (P) to the probability of the event not happening (1-P). It's a way of expressing the likelihood of an event. If the odds are greater than 1, the event is more likely to happen than not, and vice versa.

log(odds) → LoR

→ 3 odds → $\boxed{\frac{1}{6}}$ $\boxed{\frac{5}{6}}$

odds → $\frac{1/6}{5/6} = \boxed{\frac{1}{5}} = 0.2$ $\boxed{1 \text{ in } 5}$ odds

$$\text{Odds} = \frac{P}{1-p}$$

p → prob of success

$\boxed{0-1}$

✓✓✓✓✗ → $\frac{4/5}{1/5}$ $\boxed{\frac{4}{1}} \rightarrow 4$

→ $\boxed{0 \rightarrow \infty}$ →

✗✗✗✗✗
✗✗✗✗✗ → $\frac{1/5}{4/5} = \frac{1}{4} \rightarrow 0.25$ $\frac{1}{9} =$ 0

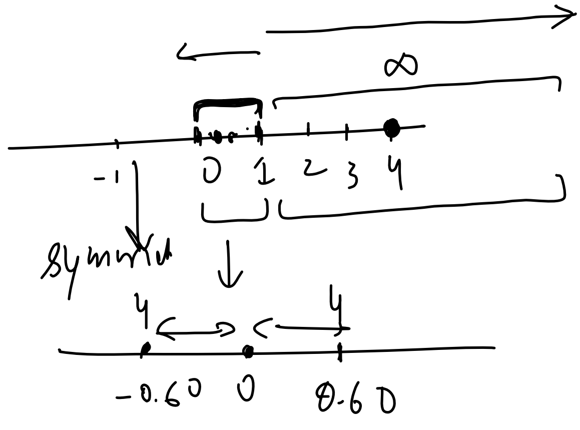
✓✓✓✓✗ → $\frac{4}{1}$ 39 399 3999

log(odds) $\log_e(\text{odds})$ →

odds

✗✓✓✓✓ → $\frac{4}{1} \rightarrow \boxed{4}$ →

✓✗✗✗✗ → $\frac{1}{4}$ 0.25



Another Interpretation of Logistic Regression

05 July 2023 07:21

log(odds) → LOR

$$\log(\text{odds}) \Rightarrow \log\left(\frac{p}{1-p}\right)$$

$\beta_0 = 0 \quad \beta_1 = 1 \quad \beta_2 = 2$

cgpa	iq	placement	$\boxed{\hat{y}} \to p$
8	80	y	
—	—	1	0.93
—	—	0	0.37
—	—	0	0.21

$p(1) \to p$

$$\hat{y} = p(1) = p = \boxed{\frac{1}{1+e^{-\beta x}}}$$

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2$$

$$\boxed{0 + 1 \times 8 + 2 \times 80}$$

$$p = \frac{1}{1+e^{-\beta x}} \Rightarrow 1 + e^{-\beta x} = \frac{1}{p}$$

$$e^{-\beta x} = \frac{1}{p} - 1 \Rightarrow e^{-\beta x} = \frac{1-p}{p}$$

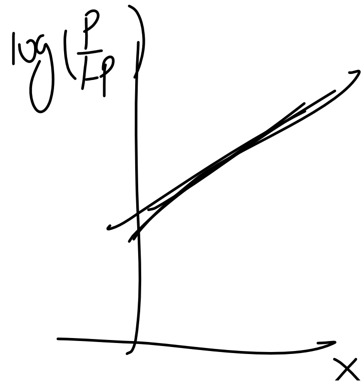
$$\frac{1}{e^{\beta x}} = \frac{1-p}{p} \Rightarrow \frac{p}{1-p} = e^{\beta x}$$

$$\log\left(\frac{p}{1-p}\right) = \log\left(e^{\beta x}\right)$$

$$\log\left(\frac{p}{1-p}\right) = \beta x$$

$$\boxed{\log\left(\frac{p}{1-p}\right)} = \beta X$$

log odds

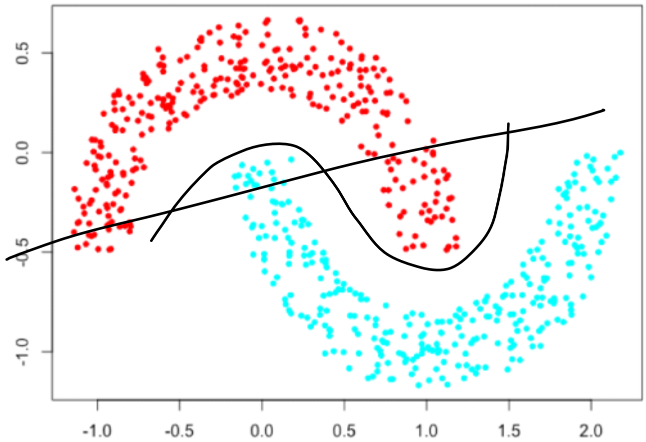


$$\underbrace{\log\left(\frac{\hat{y}}{1-\hat{y}}\right) \rightarrow x}_{\text{linear}}$$

$$\underbrace{\hat{y} \rightarrow x}_{\text{linear}}$$

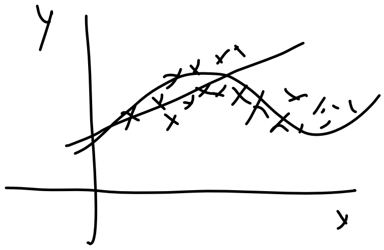
Polynomial Features

05 July 2023 10:18



deg → ince → over

small → underfit



Polynomial features $\boxed{X}$

→ X → degree → 2

β_0 β_1

X^0 | X^1 | X^2

↑ ↑ ↑

β_1 β_2 β_3

β_0

→ draw non-linear decision boun

↓

x_1 x_2 degree=2

x_1^0 x_1^1 x_1^2 x_2^0 x_2^1 x_2^2

Regularization in Logistic Regression

05 July 2023 07:22

train – wm ↓
test ↑

Regularization is a technique used in machine learning models to prevent overfitting, which occurs when a model learns the noise along with the underlying pattern in the training data. Overfitting leads to poor generalization performance when the model is exposed to unseen data.

In the context of linear models like linear regression and logistic regression, regularization works by adding a penalty term to the loss function that the model tries to minimize. This penalty term discourages the model from assigning too much importance to any single feature, which helps to prevent overfitting.

The most common types of regularization in linear models are L1 and L2 regularization:

1. L1 regularization (Lasso Regression): This technique adds a penalty term equal to the absolute value of the magnitude of the coefficients. Mathematically, it's represented as the sum of the absolute values of the weights (||w||_1). This can lead to sparse models, where some feature weights can become exactly zero. This property makes L1 regularization useful for feature selection.
2. L2 regularization (Ridge Regression): This technique adds a penalty term equal to the square of the magnitude of the coefficients. Mathematically, it's represented as the sum of the squared values of the weights (||w||_2^2). L2 regularization tends to spread the weight values more evenly across features, leading to smaller, but non-zero, weights.

mse

$$\sum (y_i - \hat{y}_i)^2 + \lambda \|w\|^2$$

There's also Elastic Net regularization, which is a combination of L1 and L2 regularization. The contribution of each type can be controlled with a separate hyperparameter.

In all these techniques, the amount of regularization to apply is controlled by a hyperparameter, often denoted as λ (lambda). Higher values of λ mean more regularization, leading to simpler models that might underfit the data. Lower values of λ mean less regularization, leading to more complex models that might overfit the data. The optimal value of λ is typically found through cross-validation.

λ → hyper

λ increase → underfit

λ small → overfit

λ → bias variance trade

$$L = -\frac{1}{n}\sum_{i=1}^{n} y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i) + \lambda\|\beta\|^2 \quad \text{L2 reg}$$

$$+ \lambda\|\beta\| \rightarrow \text{L1 reg}$$

$$C = \frac{1}{\lambda}$$

Hyperparameters

05 July 2023 10:18

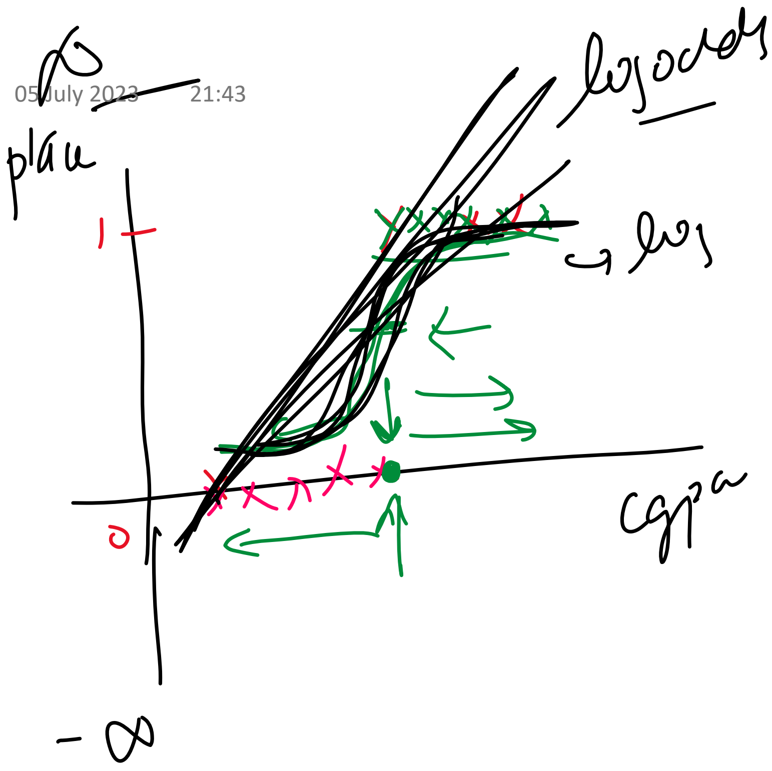
Tasks

05 July 2023 16:28

1. Find the solution for regularized loss function
2. Apply hyper parameter tuning to a real world dataset

analyse → world's best

05 July 2023 21:43



cgpa	placement
8	1
6	1
4	0
9	1
⋮