From Regression to Classification

Hao Ni
University College London
The Alan Turing Institute

Classification

Classification

Classification is one type of supervised learning where the output is categorical.

Categorical Variable

A categorical variable means that the variable has only finite many possible values of the output, denoted by $\mathcal{Y}$. W.l.o.g $\mathcal{Y} = \{1, \cdots, n_o\}$, where n_o denotes the number of possible categories.

Regression v.s. Classification

The main difference between classification and regression is the type of output variable.

- Regression: Continuous output ($\mathcal{Y} = \mathbb{R}^d$).
- Classification: Categorical output ($\mathcal{Y}$ is a finite set, i.e. $\mathcal{Y} = \{1, \cdots, n_o\}$, where n_o denotes the number of possible categories.)
  1. binary classification ($n_o = 2$).
  2. multi-class classification ($n_o > 2$).

Categorical Variables and Representation

Two representations of the categorical variable

- Integer encoding (the i^{th} class is represented using an integer i);
- One-hot vector encoding (the i^{th} class is represented using a binary vector of length n_o, which has the unique non-zero element at the i^{th} position).

Example

The representation of the price movement direction are given below:

Price Movement Direction	up	down	no change
Integer encoding	1	2	3
One-hot encoding	100	010	001

Models

Remark

The expectation of a random categorical variable does not make sense!! Hence the main objective of the classification is to estimate *the probability of the output being y on conditional on an input x*.

Models

In classification framework, a model $f_\theta : \mathcal{X} \to \mathbb{R}^{n_o}$ is to approximate the conditional probability of the output being y given input x, in formula,

$$\langle f_\theta(x), \bar{y} \rangle \approx \mathbb{P}[y|x],$$

where $\bar{y}$ is one-hot encoding of the class y and $\langle ., . \rangle$ is the inner product of two vectors of length n_o.

Loss Function

Cross Entropy

For discrete probability distributions p and q with the same support $\mathcal{Y}$, the cross entropy is defined to be

$$H(p, q) := -\sum_{j \in \mathcal{Y}} p(j) \log(q(j)).$$

Definition (Cross Entropy Loss Function)

The cross entropy loss function $Q_\theta : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is defined to be

$$Q_\theta(x, y) = -\langle \bar{y}, \log f_\theta(x) \rangle,$$

where $x \in \mathcal{X}$, $\bar{y}$ is one-hot encoding in $\mathcal{Y}$, θ are model parameters of f_θ, and $\langle ., . \rangle$ is the inner product.

Empirical Cross Entropy Loss

The empirical cross entropy loss function is given as the average of the above cross entropy loss function evaluated at all samples:

$$L(\theta|\mathcal{D}) = -\frac{1}{N}\sum_{i=1}^{N}\langle y_i, \log f_\theta(x_i)\rangle.$$

Maximum Likelihood Estimation (MLE) Interpretation of Cross Entropy

$$\prod_{i=1}^{N}\langle f_\theta(x_i), \bar{y}_i\rangle \to \max,$$

which is equivalent to minimize the negative log-likelihood ratio, i.e.

$$-\sum_{i=1}^{N}(\langle \log f_\theta(x_i), \bar{y}_i\rangle) \to \min.$$

Prediction

Let θ_* denote the optimal model parameter. Then for any new given input x_*, the estimated output class is given by the class i, which achieves the highest estimated probability, i.e.

$$\hat{y}_* = \arg\max_{i \in \mathcal{Y}} f_{\theta_*}^{(i)}(x_*),$$

where $f_{\theta_*}^{(i)}(x_*)$ is the i^{th} coordinate of $f_{\theta_*}(x_*)$.

Example (Prediction of Multi-class Classification)

class index	1	2	3	4
predicted prob.	0.1	0.2	0.6	0.1

$\downarrow$

predicted label $= 3$

Imbalanced classification

Imbalanced classification refers to the classification tasks, where the distribution of output class is far from the even distribution. [1]

Remark

The prediction of the imbalance classification might be different from the method mentioned in the above.

Conclusion

Summary

In this video, we cover

- what is the classification?
- How to extend the regression framework to classification, including the model, loss function and prediction.

We will follow by discussing the test metric of the classification in the next step.

Thanks for your attention!

References I

Aida Ali, Siti Mariyam Shamsuddin, and Anca L Ralescu.
Classification with class imbalance problem: A review.
Int. J. Advance Soft Compu. Appl, 7(3), 2015.