Monchwe Mpundo

201726904

NAML 841

26 Jul 2021

SECTION A

1.1 Given data $D = \{(x, y)\}^n$, where n is the number of data points

From the Bayesian perspective

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

- $p(\theta, y) = p(\theta|y)p(y)$

$$p(\theta, y) = p(y|\theta)p(\theta)$$
$$p(\theta|y)p(y) = p(y|\theta)p(\theta)$$

$p(\theta) \to$ prior

$p(y|\theta) \to$ Likelihood

$p(\theta|y) \to$ Posterior

$P(y) \to$ Marginal likelihood

Bayesian linear regression

$$p(y, x|\theta)$$
$$p(y, x, |\theta) = p(y|x, \theta)p(x|\theta)$$

- Assumption that x does not depend on parameters θ.

given in test data to predict y, $p(y, x|\theta) = p(y|x, \theta)p(x)$

- probability of data $p(x) \to$ normalising const.

$$p(y, x|\theta) \propto p(y|x, \theta)$$

$$y = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \ldots \theta_n x_n^{(i)} + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2)$$

1.2 We can make assumption about OLS regression

- random distribution Sampling of observation
- The conditional mean should be zero

Therefore,

Assume Gaussian likelihood: $y_i \sim N(\mu, \sigma^2)$

2.1 If a function that models the relationship between x and y is $\hat{y} = f(x, w)$

- w_0 and w_1 for learning function
- w_2 for amplitude
- w_3 for vertical movement

$$f(x; w) = w_3 + \frac{w_2}{1 + e^{-(w_0 + w_1 x)}}$$

3.1 Let $x \in [-5, 5]$

4.1 Weaknesses of using k-means

- Transformations clusters are non-circular and these circular clusters are of poor fit. This result in mixing of cluster assignments where resulting circles overlap
- It lacks flexibility in cluster shape
- Lacks probabilistic cluster assignment.

4.2 GMM contains a probabilistic model

- measures that probability that any point belongs to the given cluster using the predict_proba method
- Uses expectation-maximation approach