CS464 Introduction to Machine Learning Spring 2022 Homework 1

Enver Yiğitler

March 11, 2022

1 The Online Shopping Case

Question 1.1

$$\mathbb{P}(F_P|P) = \frac{95}{100} \quad \mathbb{P}(F_P|M) = \frac{60}{100} \quad \mathbb{P}(F_P|U) = \frac{10}{100}$$
$$\mathbb{P}(P) = \frac{45}{100} \quad \mathbb{P}(M) = \frac{30}{100} \quad \mathbb{P}(U) = \frac{25}{100}$$

$$\mathbb{P}(F_P) = \sum_i \mathbb{P}(F_P|X_i)\mathbb{P}(X_i) = \mathbb{P}(F_P|P)\mathbb{P}(P) + \mathbb{P}(F_P|M)\mathbb{P}(M) + \mathbb{P}(F_P|U)\mathbb{P}(U)$$
$$= 0.95 \cdot 0.45 + 0.6 \cdot 0.3 + 0.1 \cdot 0.25 = 0.6325$$

Question 1.2

$$\mathbb{P}(P|F_P) = \frac{\mathbb{P}(F_P|P) \cdot \mathbb{P}(P)}{\sum_i \mathbb{P}(F_P|X_i)\mathbb{P}(X_i)}$$

$$= \frac{0.4275}{0.6325} = 0.6758$$

Question 1.3

$$\mathbb{P}(P|F_N) = \frac{\mathbb{P}(F_N|P) \cdot \mathbb{P}(P)}{\sum_i \mathbb{P}(F_N|X_i)\mathbb{P}(X_i)}$$

$$= \frac{0.0225}{0.05 \cdot 0.45 + \ 0.4 \cdot 0.3 + 0.9 \cdot 0.25} = 0.061$$

2 MLE and MAP

Question 2.1

$$\mathcal{N}(X|\mu,\sigma) = \Big(\frac{1}{\sigma\sqrt{2\pi}}\Big)^N \prod_{i=1}^{N} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}}$$

$$\mu_{MLE} = \underset{\mu}{\operatorname{argmax}}\ \mathcal{N}(X|\mu,\sigma)$$

Log-likelihood:

$$\begin{aligned}\ln\mathcal{N}(X|\mu,\sigma) &= \ln\Big[\Big(\frac{1}{\sigma\sqrt{2\pi}}\Big)^N \prod_{i=1}^{N} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}}\Big] \\ &= -N\ln\sigma\sqrt{2\pi} - \sum_{i=1}^{N}\frac{(x_i-\mu)^2}{2\sigma^2}\end{aligned}$$

We can set derivative of log-likelihood to zero to find the maximum:

$$\begin{aligned}\frac{d}{d\mu}\ln\mathcal{N}(X|\mu,\sigma) &= \frac{d}{d\mu}\Big[\sigma\sqrt{2\pi} - \sum_{i=1}^{N}\frac{(x_i-\mu)^2}{2\sigma^2}\Big] \\ &= \sum_{i=1}^{N}\frac{(x_i-\mu)}{\sigma^2} = 0 \\ &= \sum_{i=1}^{N} x_i - N\mu = 0 \\ \mu_{MLE} &= \frac{1}{N}\sum_{i=1}^{N} x_i\end{aligned}$$

Question 2.2

$$\mu \sim \mathcal{N}(\nu, \alpha^2)$$

$$P(\mu | X) \propto P(X|\mu)\, P(\mu)$$

$$= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}} \times \frac{1}{\sqrt{2\pi\alpha}} e^{-\frac{(\mu-\nu)^2}{2\alpha^2}}$$

$$\ln P(\mu|X) \propto \sum_{i=1}^{N} \left(\frac{x_i - \mu}{\sigma}\right)^2 + \left(\frac{\mu-\nu}{\alpha}\right)^2$$

$$\frac{d}{d\mu} \ln P(\mu|X) \propto -\frac{2}{\sigma^2}\left(\sum_{i}^{N} x_i - N\mu\right) + \frac{2}{\alpha^2}(\mu - \nu) = 0$$

After solving for μ we find

$$\mu_{MAP} = \frac{\alpha^2 N}{\alpha^2 N + \sigma^2}\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right) + \frac{\sigma^2}{\alpha^2 N + \sigma^2} \nu$$

Question 2.3

We found

$$\mu_{MAP} = \frac{\alpha^2 N}{\alpha^2 N + \sigma^2} \left(\frac{1}{N} \sum_{i=1}^{N} x_i \right) + \frac{\sigma^2}{\alpha^2 N + \sigma^2} \nu$$

$$\lim_{N \to \infty} \mu_{MAP} = \frac{1}{1 + \underbrace{\frac{\sigma^2}{\alpha^2 N}}_{0}} \left(\frac{1}{N} \sum_{i=1}^{N} x_i \right) + \cancel{\frac{\sigma^2}{\alpha^2 N + \sigma^2} \nu} \; 0$$

$$= \frac{1}{N} \sum_{i=1}^{N} x_i = \mu_{MLE}$$

Question 3.1

1. %28.5

2. Data is skewed towards normal mail samples. When we have class imbalance, Naive Bayes classifier tends to be more biased towards the majority class. We can change performance metrics. Accuracy is not a right choice for imbalanced data. If we can somehow generate data for minority class this would be the best solution.

3. If the classifier randomly assign classes for an imbalanced dataset, this would lead to a good accuracy score due to the high percentage of majority class. So, accuracy is misleading for unbalanced datasets.

Question 3.2 Multinomial Naive Bayes Model

Results:
TP: 289 FP: 15
FN: 28 TN: 703

Accuracy: 0.9584541062801932
Recall: 0.9116719242902208
Precision: 0.9506578947368421
F-measure: 0.9307568438003221
Wrong predictions: 43

Question 3.3 Multinomial Naive Bayes Mode with Dirichlet prior

Results:
alpha = 5
TP: 300 FP: 37
FN: 17 TN: 681

Accuracy: 0.9478260869565217
Recall: 0.9463722397476341
Precision: 0.8902077151335311
F-measure: 0.9174311926605505
Wrong predictions: 54

for alpha = 2, Accuracy: 0.9487922705314009
for alpha = 4, Accuracy: 0.9478260869565217
for alpha = 10, Accuracy: 0.9458937198067633
for alpha = 30, Accuracy: 0.9391304347826087
for alpha = 100, Accuracy: 0.8753623188405797.

Question 3.4 Bernoulli Naive Bayes Model

Results:
TP: 280 FP: 15
FN: 37 TN: 703

Accuracy: 0.9497584541062802
Recall: 0.8832807570977917
Precision: 0.9491525423728814
F-measure: 0.9150326797385621
Wrong predictions: 52

Question 3.5

List of comparison of models for each metrics:

Accuracy: Multinomial Naive Bayes Mode > Bernoulli Naive Bayes Mode > Multinomial Naive Bayes Mode with Dirichlet prior

Recall: Multinomial Naive Bayes Mode with Dirichlet prior > Multinomial Naive Bayes Model > Bernoulli Naive Bayes Model

Precision: Multinomial Naive Bayes Model > Bernoulli Naive Bayes Model > Multinomial Naive Bayes Mode with Dirichlet prior

F-measure: Multinomial Naive Bayes Model > Multinomial Naive Bayes Mode with Dirichlet prior > Bernoulli Naive Bayes Model

We can see that Multinomial Naive Bayes Model is superior than other models for each metrics other than recall. The reason of inferiority of the Multinomial Naive Bayes Mode with Dirichlet prior is probably prior estimation is not correct for the parameter. Difference between Multinomial Naive Bayes and Bernoulli Naive Bayes can be explained with the extra information that Multinomial model is using. Accuracy would be misleading for this dataset due to class imbalance. F-measure gives more general assessment of the model. So we can use f-measure metric to choose the model that we will use for real world applications.