

Missing values and Naive Bayes

0.1 A modification of Naive Bayes to deal with missing values

- Training

When we fit $P(x_i|y)$ for feature x_i , we can just use all available values and ignore missing values.

- Testing[?]

If a test data point has some missing features, say x_1 , we can marginalizing it out.

$$\begin{aligned} P(y|x_2, \dots, x_d) &\propto P(y)P(x_2, \dots, x_d|y) \\ &= P(y) \sum_{x_1} P(x_1, x_2, \dots, x_d|y) \\ &= P(y) \sum_{x_1} \prod_i P(x_i|y) \\ &= P(y) \left(\sum_{x_1} P(x_1|y) \right) \prod_{i=2,3,\dots,d} P(x_i|y) \\ &= P(y) \prod_{i=2,3,\dots,d} P(x_i|y) \end{aligned}$$

So the classification rule essentially ignores the missing feature x_1 , and uses other available feature values.

References

- [1] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.
- [2] Maytal Saar-Tsechansky and Foster Provost. Handling missing values when applying classification models. *Journal of Machine Learning Research*, 2007.