Machine Learning

Performance - Naive Bayes

Edgar Francisco Roman-Rangel
edgar.roman@itam.mx

Digital Systems Department.
Instituto Tecnológico Autónomo de México, ITAM.

Diplomado - UAEM

May 1^{st}, 2021.

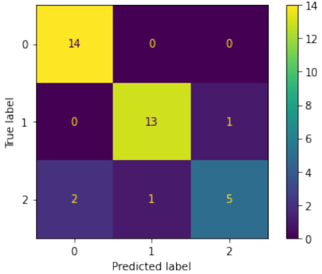
Outline

Performance metrics

Naive Bayes Classifier

Confusion matrix

		Actual Value (as confirmed by experiment)	
		positives	negatives
Predicted Value (predicted by the test)	positives	**TP** True Positive	**FP** False Positive
	negatives	**FN** False Negative	**TN** True Negative



Other metrics, I

Accuracy:

$$a = \frac{TP + TN}{TP + FP + FN + TN}.$$

Precision:

$$p = \frac{TP}{TP + FP}.$$

Sensitivity (recall):

$$r = \frac{TP}{TP + FN}.$$

Specificity (selectivity):

$$s = \frac{TN}{TN + FP}.$$

Other metrics, II

False negative rate (FNR):

$$FNR = \frac{FN}{TP + FN}.$$

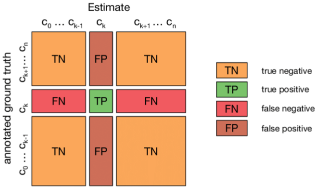
False positive rate (FPR):

$$FPR = \frac{FP}{FP + TN}.$$

F1 score:

$$F1 = \frac{2TP}{2TP + FP + FN}.$$

Confusion matrix



	Actual Dog	Actual Cat	Actual Rabbit
Classified Dog	23	12	7
Classified Cat	11	29	13
Classified Rabbit	4	10	24

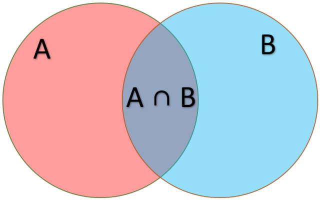
Outline

Performance metrics

Naive Bayes Classifier

Bayes Theorem

$$p(A, B) = p(A)p(B|A).$$



$$p(A)p(B|A) = p(B)p(A|B),$$
$$p(B|A) = \frac{p(B)p(A|B)}{p(A)}.$$

Bayes estimator

Classification method based on the Bayes Theorem:

$$p(y|\mathbf{x}) = \frac{p(\mathbf{x}|y)p(y)}{p(\mathbf{x})},$$

- $p(y|\mathbf{x})$: **posterior** (what we are looking for).
  Probability of class y, given input $\mathbf{x}$.
- $p(\mathbf{x}|y)$: **prior** (observation from data).
  Probability of observing input $\mathbf{x}$ when data point is of class y.
- $p(y)$: **likelihood**.
  Probability of class y in our data set.
- $p(\mathbf{x})$: **evidence**.
  Probability of input pattern $\mathbf{x}$ in our data set.

Naive Bayes Classifier

- Assume independence among features (elements of vector $\mathbf{x}$).

$$p(\mathbf{x}) = p(x_1)p(x_2)\dots p(x_N).$$

  This is rare in real world scenarios. However, it does work in practice.
- Evidence $p(\mathbf{x})$ is the same for a fixed data set.

Therefore, the naive version becomes:

$$\begin{aligned} p(y|\mathbf{x}) &\propto p(\mathbf{x}|y)p(y), \\ &\propto \prod_{n=1}^{N} p(x_n|y)p(y). \end{aligned}$$

Univariate example

Probability of 'dog' (d) if there are '4 legs'.

$$p(y = \mathsf{d}|\mathsf{4}) = \frac{p(\mathsf{4}|\mathsf{d})p(\mathsf{d})}{p(\mathsf{4})},$$

suppose from our data we count:

- $p(\mathsf{4}|\mathsf{d}) = 4/5$.
- $p(\mathsf{d}) = 2/3$.
- $p(\mathsf{4}) = 1/10$.

$$p(y = \mathsf{d}|\mathsf{4}) = \frac{4/5 \cdot 2/3}{1/10} = \frac{8/15}{1/10} = 5.3$$

Let's say it is a dog if all other $p(y|\mathsf{4})$ are less than 5.3.

Multivariate example

I love biking. Should I go biking today?

Let us use:

- $\mathbf{x} = [x_1 = \text{sky}, x_2 = \text{temperature}, x_3 = \text{wind}]$.
- $y = \{0, 1\}$ (biking or not biking).

n	sky	temp	wind	biking
1	sunny	hot	FALSE	0
2	sunny	hot	TRUE	0
3	cloudy	hot	FALSE	1
4	rainy	mild	FALSE	1
5	rainy	cool	FALSE	1
6	rainy	cool	TRUE	0
7	cloudy	cool	TRUE	1
8	sky	mild	FALSE	0
9	sky	cool	FALSE	1
10	rainy	mild	FALSE	1
11	sky	mild	TRUE	1
12	cloudy	mild	TRUE	1
13	cloudy	hot	FALSE	1
14	rainy	mild	TRUE	0

Notes on Naive Bayes Classifier

- This method exploits probabilities.
- Easy and fast.
- Performs better than other methods (assuming independence).
- Zero frequencies might be problematic.

Used for:

- Credit analysis.
- Spam detector.
- Medical analysis.
- Recommendation systems.

Q&A

Thank you!

edgar.roman@itam.mx