

Machine Learning

Prepared by Dr. Yasser Abdelhamid

DS 620
Machine Learning

Lecture 6





OUTLINE

- Classification
- Naïve Bayes Algorithm
- Measuring accuracy

CLASSIFICATION



CLASSIFICATION

❖ Classification is the process of dividing up objects so that each is assigned to one of a number of mutually exclusive and exhaustive categories known as classes.

❖ Examples:

- Buying a product.
- Having a disease.
- Stock price.
- ...

MUTUALLY EXCLUSIVE CLASSES

- ❖ Mutually exclusive classes, in the context of classification, refer to classes or categories that do not overlap or have any elements in common.
- ❖ If two classes are mutually exclusive, it means that an instance can only belong to one class and cannot belong to multiple classes simultaneously.

EXHAUSTIVE CLASSES

- ❖ Exhaustive classes refer to a situation where all possible categories or classes are considered and included in the classification scheme.
- ❖ It means that every instance or observation must be assigned to one of the predefined categories without any instance being left unclassified.

NAÏVE BAYES CLASSIFIERS

NAÏVE BAYES CLASSIFICATION

- ❖ Naive Bayes is a classification algorithm that is based on Bayes' theorem.
- ❖ It is a simple yet effective probabilistic algorithm widely used for text classification, spam filtering, sentiment analysis, and other machine learning tasks.

BASICS OF PROBABILITY THEORY

- ❖ Given a dataset that consists of **n** instances.
- ❖ $P(E)$ is the probability of an event **E** that has occurred **n(E)** times, is calculated as follows:

$$P(E) = \frac{n(E)}{n}$$

CONDITIONAL AND JOINT PROBABILITY

❖ Conditional probability = $P(A|B) = \frac{n(A,B)}{n(B)}$

❖ Joint probability = $P(A,B) = P(A|B)P(B)$

EXAMPLE

❖ Let's have the shown dataset of Alexandria train arrival time.

❖ Possible events (classes) are:

- E1 = cancelled = 0.05
- E2 = very late = 0.10
- E3 = late = 0.15
- E4 = on-time = 0.70

day	season	wind	rain	class
weekday	spring	none	none	on time
weekday	winter	none	slight	on time
weekday	winter	none	slight	on time
weekday	winter	high	heavy	late
saturday	summer	normal	none	on time
weekday	autumn	normal	none	very late
holiday	summer	high	slight	on time
sunday	summer	normal	none	on time
weekday	winter	high	heavy	very late
weekday	summer	none	slight	on time
saturday	spring	high	heavy	cancelled
weekday	summer	high	slight	on time
saturday	winter	normal	none	late
weekday	summer	high	none	on time
weekday	winter	normal	heavy	very late
saturday	autumn	high	slight	on time
weekday	autumn	none	heavy	on time
holiday	spring	normal	slight	on time
weekday	spring	normal	none	on time
weekday	spring	normal	slight	on time

EXAMPLE

❖ How should we use probabilities to find the most likely classification for an unseen instance such as the one below?

weekday	winter	high	heavy	????
---------	--------	------	-------	------

day	season	wind	rain	class
weekday	spring	none	none	on time
weekday	winter	none	slight	on time
weekday	winter	none	slight	on time
weekday	winter	high	heavy	late
saturday	summer	normal	none	on time
weekday	autumn	normal	none	very late
holiday	summer	high	slight	on time
sunday	summer	normal	none	on time
weekday	winter	high	heavy	very late
weekday	summer	none	slight	on time
saturday	spring	high	heavy	cancelled
weekday	summer	high	slight	on time
saturday	winter	normal	none	late
weekday	summer	high	none	on time
weekday	winter	normal	heavy	very late
saturday	autumn	high	slight	on time
weekday	autumn	none	heavy	on time
holiday	spring	normal	slight	on time
weekday	spring	normal	none	on time
weekday	spring	normal	slight	on time

THE EASY WAY

❖ Using only the **frequency** of classes, we can expect that the train will arrive on time.

❖ Is that satisfying?

day	season	wind	rain	class
weekday	spring	none	none	on time
weekday	winter	none	slight	on time
weekday	winter	none	slight	on time
weekday	winter	high	heavy	late
saturday	summer	normal	none	on time
weekday	autumn	normal	none	very late
holiday	summer	high	slight	on time
sunday	summer	normal	none	on time
weekday	winter	high	heavy	very late
weekday	summer	none	slight	on time
saturday	spring	high	heavy	cancelled
weekday	summer	high	slight	on time
saturday	winter	normal	none	late
weekday	summer	high	none	on time
weekday	winter	normal	heavy	very late
saturday	autumn	high	slight	on time
weekday	autumn	none	heavy	on time
holiday	spring	normal	slight	on time
weekday	spring	normal	none	on time
weekday	spring	normal	slight	on time

CONDITIONAL PROBABILITY

❖ What is the probability of the train being on time if we know that the season is winter?

day	season	wind	rain	class
weekday	spring	none	none	on time
weekday	winter	none	slight	on time
weekday	winter	none	slight	on time
weekday	winter	high	heavy	late
saturday	summer	normal	none	on time
weekday	autumn	normal	none	very late
holiday	summer	high	slight	on time
sunday	summer	normal	none	on time
weekday	winter	high	heavy	very late
weekday	summer	none	slight	on time
saturday	spring	high	heavy	cancelled
weekday	summer	high	slight	on time
saturday	winter	normal	none	late
weekday	summer	high	none	on time
weekday	winter	normal	heavy	very late
saturday	autumn	high	slight	on time
weekday	autumn	none	heavy	on time
holiday	spring	normal	slight	on time
weekday	spring	normal	none	on time
weekday	spring	normal	slight	on time

CONDITIONAL PROBABILITY

- ❖ We can calculate this as the number of times class = **on time (A)** and season = **winter** (in the same instance),
- ❖ Divide it by the number of times the season is winter(**B**), which is $2/6 = 0.33$.

$$P(A|B) = \frac{n(A,B)}{n(B)}$$

- ❖ $P(\text{class} = \text{on time} \mid \text{season} = \text{winter})$.

day	season	wind	rain	class
weekday	spring	none	none	on time
weekday	winter	none	slight	on time
weekday	winter	none	slight	on time
weekday	winter	high	heavy	late
saturday	summer	normal	none	on time
weekday	autumn	normal	none	very late
holiday	summer	high	slight	on time
sunday	summer	normal	none	on time
weekday	winter	high	heavy	very late
weekday	summer	none	slight	on time
saturday	spring	high	heavy	cancelled
weekday	summer	high	slight	on time
saturday	winter	normal	none	late
weekday	summer	high	none	on time
weekday	winter	normal	heavy	very late
saturday	autumn	high	slight	on time
weekday	autumn	none	heavy	on time
holiday	spring	normal	slight	on time
weekday	spring	normal	none	on time
weekday	spring	normal	slight	on time

PRIOR PROBABILITY

- ❖ Probability that is estimated before any other information is available is called **prior probability**.

POSTERIOR PROBABILITY

- ❖ The probability that we can calculate for the classification after we have obtained an information is called **posterior probability**.

MOST LIKELY CLASSIFICATION (SINGLE ATTRIBUTE)

- ❖ What is the **most likely classification** of an instance given that rain = heavy?
- ❖ We calculate the posterior probability for each class and select the highest probability.

MOST LIKELY CLASSIFICATION (SINGLE ATTRIBUTE)

$$P(\text{class} = \text{on time} \mid \text{rain} = \text{heavy}) = 1/5 = 0.2$$

$$P(\text{class} = \text{late} \mid \text{rain} = \text{heavy}) = 1/5 = 0.2$$

$$P(\text{class} = \text{very late} \mid \text{rain} = \text{heavy}) = 2/5 = 0.4$$

$$P(\text{class} = \text{cancelled} \mid \text{rain} = \text{heavy}) = 1/5 = 0.2$$

❖ So we can conclude that the most likely class is “**very late**”.

day	season	wind	rain	class
weekday	spring	none	none	on time
weekday	winter	none	slight	on time
weekday	winter	none	slight	on time
weekday	winter	high	heavy	late
saturday	summer	normal	none	on time
weekday	autumn	normal	none	very late
holiday	summer	high	slight	on time
sunday	summer	normal	none	on time
weekday	winter	high	heavy	very late
weekday	summer	none	slight	on time
saturday	spring	high	heavy	cancelled
weekday	summer	high	slight	on time
saturday	winter	normal	none	late
weekday	summer	high	none	on time
weekday	winter	normal	heavy	very late
saturday	autumn	high	slight	on time
weekday	autumn	none	heavy	on time
holiday	spring	normal	slight	on time
weekday	spring	normal	none	on time
weekday	spring	normal	slight	on time

NAÏVE BAYES THEOREM

- ❖ The Naïve Bayes algorithm gives us a way of combining the prior probability and conditional probabilities in a single formula, which **we can use to calculate the probability of each of the possible classifications in turn.**

NAÏVE BAYES THEOREM

- ❖ Having done this we choose the classification with the **largest value**.
- ❖ The meaning of the word |"Naïve" is that the effect of the value of one attribute on the probability of a given classification is **independent** of the values of the other attributes.

NAÏVE BAYES THEOREM

❖ Given that $P(A, B) = P(A|B)P(B)$

❖ $P(B, A) = P(B|A)P(A)$

❖ Using symmetry rule $P(A, B) = P(B, A)$

❖ Then $P(B|A)P(A) = P(A|B)P(B)$

❖ We can conclude with Bayes theorem which is:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

❖ i.e. $Posterior = \frac{Likelihood \times Prior}{Evidence}$

NAÏVE BAYES CLASSIFICATION

- ❖ Given a set of k mutually exclusive and exhaustive classifications $c_1, c_2, \dots, c_k$,
- ❖ Each class c_i has prior probability $P(c_i)$, and
- ❖ Given n attributes $a_1, a_2, \dots, a_n$ which for a given instance have values $v_1, v_2, \dots, v_n$ respectively.

NAÏVE BAYES CLASSIFICATION

- ❖ The posterior probability of class c_i occurring for the specified instance can be shown to be proportional to:

$$P(c_i) \times P(a_1 = v_1 \text{ and } a_2 = v_2 \dots \text{ and } a_n = v_n \mid c_i)$$

NAÏVE BAYES CLASSIFICATION

- ❖ Making the assumption that the attributes are **independent**, the value of this expression can be calculated using the product:

$$P(c_i) \times P(a_1 = v_1 \mid c_i) \times P(a_2 = v_2 \mid c_i) \times \dots \times P(a_n = v_n \mid c_i)$$

- ❖ Which is $P(c_i) \times \prod_{j=1}^n P(a_j = v_j \mid \text{class} = c_i)$.
- ❖ We calculate this product for each value of **i** from **1** to **k** and choose the classification that has the largest value.

day	season	wind	rain	class
weekday	spring	none	none	on time
weekday	winter	none	slight	on time
weekday	winter	none	slight	on time
weekday	winter	high	heavy	late
saturday	summer	normal	none	on time
weekday	autumn	normal	none	very late
holiday	summer	high	slight	on time
sunday	summer	normal	none	on time
weekday	winter	high	heavy	very late
weekday	summer	none	slight	on time
saturday	spring	high	heavy	cancelled
weekday	summer	high	slight	on time
saturday	winter	normal	none	late
weekday	summer	high	none	on time
weekday	winter	normal	heavy	very late
saturday	autumn	high	slight	on time
weekday	autumn	none	heavy	on time
holiday	spring	normal	slight	on time
weekday	spring	normal	none	on time
weekday	spring	normal	slight	on time

Classes

Day

Season

Wind

Rain

	class = on time	class = late	class = very late	class = cancelled
day = weekday	9/14 = 0.64	1/2 = 0.5	3/3 = 1	0/1 = 0
day = saturday	2/14 = 0.14	1/2 = 0.5	0/3 = 0	1/1 = 1
day = sunday	1/14 = 0.07	0/2 = 0	0/3 = 0	0/1 = 0
day = holiday	2/14 = 0.14	0/2 = 0	0/3 = 0	0/1 = 0
season = spring	4/14 = 0.29	0/2 = 0	0/3 = 0	1/1 = 1
season = summer	6/14 = 0.43	0/2 = 0	0/3 = 0	0/1 = 0
season = autumn	2/14 = 0.14	0/2 = 0	1/3 = 0.33	0/1 = 0
season = winter	2/14 = 0.14	2/2 = 1	2/3 = 0.67	0/1 = 0
wind = none	5/14 = 0.36	0/2 = 0	0/3 = 0	0/1 = 0
wind = high	4/14 = 0.29	1/2 = 0.5	1/3 = 0.33	1/1 = 1
wind = normal	5/14 = 0.36	1/2 = 0.5	2/3 = 0.67	0/1 = 0
rain = none	5/14 = 0.36	1/2 = 0.5	1/3 = 0.33	0/1 = 0
rain = slight	8/14 = 0.57	0/2 = 0	0/3 = 0	0/1 = 0
rain = heavy	1/14 = 0.07	1/2 = 0.5	2/3 = 0.67	1/1 = 1
Prior Probability	14/20 = 0.70	2/20 = 0.10	3/20 = 0.15	1/20 = 0.05

day	season	wind	rain	class
weekday	spring	none	none	on time
weekday	winter	none	slight	on time
weekday	winter	none	slight	on time
weekday	winter	high	heavy	late
saturday	summer	normal	none	on time
weekday	autumn	normal	none	very late
holiday	summer	high	slight	on time
sunday	summer	normal	none	on time
weekday	winter	high	heavy	very late
weekday	summer	none	slight	on time
saturday	spring	high	heavy	cancelled
weekday	summer	high	slight	on time
saturday	winter	normal	none	late
weekday	summer	high	none	on time
weekday	winter	normal	heavy	very late
saturday	autumn	high	slight	on time
weekday	autumn	none	heavy	on time
holiday	spring	normal	slight	on time
weekday	spring	normal	none	on time
weekday	spring	normal	slight	on time

PREDICTED CLASS

❖ What is the predicted class of the following instance?

weekday	winter	high	heavy	????
---------	--------	------	-------	------

day	season	wind	rain	class
weekday	spring	none	none	on time
weekday	winter	none	slight	on time
weekday	winter	none	slight	on time
weekday	winter	high	heavy	late
saturday	summer	normal	none	on time
weekday	autumn	normal	none	very late
holiday	summer	high	slight	on time
sunday	summer	normal	none	on time
weekday	winter	high	heavy	very late
weekday	summer	none	slight	on time
saturday	spring	high	heavy	cancelled
weekday	summer	high	slight	on time
saturday	winter	normal	none	late
weekday	summer	high	none	on time
weekday	winter	normal	heavy	very late
saturday	autumn	high	slight	on time
weekday	autumn	none	heavy	on time
holiday	spring	normal	slight	on time
weekday	spring	normal	none	on time
weekday	spring	normal	slight	on time

PREDICTED CLASS

❖class = on time

- $0.70 \times 0.64 \times 0.14 \times 0.29 \times 0.07 = 0.0013$

❖class = late

- $0.10 \times 0.50 \times 1.00 \times 0.50 \times 0.50 = 0.0125$

❖class = very late

- $0.15 \times 1.00 \times 0.67 \times 0.33 \times 0.67 = 0.0222$

❖class = cancelled

- $0.05 \times 0.00 \times 0.00 \times 1.00 \times 1.00 = 0.0000$

❖The largest value is for class = very late.

day	season	wind	rain	class
weekday	spring	none	none	on time
weekday	winter	none	slight	on time
weekday	winter	none	slight	on time
weekday	winter	high	heavy	late
saturday	summer	normal	none	on time
weekday	autumn	normal	none	very late
holiday	summer	high	slight	on time
sunday	summer	normal	none	on time
weekday	winter	high	heavy	very late
weekday	summer	none	slight	on time
saturday	spring	high	heavy	cancelled
weekday	summer	high	slight	on time
saturday	winter	normal	none	late
weekday	summer	high	none	on time
weekday	winter	normal	heavy	very late
saturday	autumn	high	slight	on time
weekday	autumn	none	heavy	on time
holiday	spring	normal	slight	on time
weekday	spring	normal	none	on time
weekday	spring	normal	slight	on time

MEASURING ACCURACY

SEPARATE TRAINING AND TEST SETS

- ❖ Data is split into two parts called a ***training set*** and a ***test set***.
- ❖ Training set is used to construct a classifier.
- ❖ The classifier is then used to predict the classification for the instances in the test set.
- ❖ ***Predictive accuracy*** of the classifier for the test set is $p = C/N$, where C is the correctly classified cases and N is the total number of test cases.
- ❖ This can be used as an estimate of classifier's performance on any unseen dataset.

PREDICTIVE ACCURACY

- ❖ To measure the efficiency of the classifier, predictive classifier is commonly used:

$$\text{Accuracy} = \frac{\text{Correctly classified instances}}{\text{Total number of test cases}}$$

- ❖ Either by using hold-out set method or using k-fold cross validation method.

PREDICTIVE ACCURACY

- ❖ Accuracy measure may not be efficient in some applications where classes are imbalanced.
- ❖ Example:
 - In Credit card fraud detection, 99% of the instances are classified as non-fraudulent, and only 1% are classified as fraudulent.

CONFUSION MATRIX

- ❖ A confusion matrix is an $N \times N$ matrix that describes the classification errors in detail, compared to the actual classes.
- ❖ N is the number of classes in the target attribute.

The following table displays a 2x2 confusion matrix for two classes (Positive and Negative)

Confusion Matrix		Target			
		Positive	Negative		
Model	Positive	a	b	Positive Predictive Value	$a/(a+b)$
	Negative	c	d	Negative Predictive Value	$d/(c+d)$
		Sensitivity	Specificity	Accuracy = $(a+d)/(a+b+c+d)$	
		$a/(a+c)$	$d/(b+d)$		

CONFUSION MATRIX

- ❖ **Accuracy** : The proportion of correct classifications from the overall number of cases.
- ❖ **Positive Predictive Value or Precision** : the proportion of positive cases that were correctly identified.
- ❖ **Negative Predictive Value** : the proportion of negative cases that were correctly identified.
- ❖ **Sensitivity or Recall** : the proportion of actual positive cases which are correctly identified.
- ❖ **Specificity** : the proportion of actual negative cases which are correctly identified.

Confusion Matrix		Target			
		Positive	Negative		
Model	Positive	a	b	Positive Predictive Value	$a/(a+b)$
	Negative	c	d	Negative Predictive Value	$d/(c+d)$
		Sensitivity $a/(a+c)$	Specificity $d/(b+d)$	Accuracy = $(a+d)/(a+b+c+d)$	

PRECISION AND RECALL MEASURES

- ❖ **TP (True Positive):** The number of **correct** classifications of the positive examples
- ❖ **FN (False Negative):** The number of **incorrect** classifications of the positive examples.
- ❖ **FP(False Positive):** The number of **incorrect** classifications of the negative examples.
- ❖ **TN(True Negative):** The number of **correct** classifications of the negative examples.

		Model prediction	
		Predicted Positive	Predicted Negative
Actual State	Actual positive	TP	FN
	Actual negative	FP	TN

PRECISION AND RECALL MEASURES

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{Recall} = \frac{TP}{TP + FN}$$

- ❖ Precision: is the number of correctly classified positive examples divided by the **total number of examples that are classified as positive**.
- ❖ Recall: is the number of correctly classified positive examples divided by the **total number of actual positive examples in the test set**.

		Model prediction	
		Predicted Positive	Predicted Negative
Actual State	Actual positive	TP	FN
	Actual negative	FP	TN

F1-SCORE

- ❖ F1 score combines precision and recall into one measure.

$$F_1 = \frac{2pr}{p+r}$$

- ❖ For F_1 -value to be large, both p and r must be large.



SUMMARY

- Classification
- Naïve Bayes Algorithm
- Measuring accuracy
 - Predictive accuracy
 - Confusion matrix
 - Precision and recall
 - F1 score

