Artificial Intelligence and Machine Learning
Lecture #6 part 2

Amin Noroozi
University of Wolverhampton
a.noroozifakhabi@wlv.ac.uk
https://www.linkedin.com/in/amin-n-148350218/

Classification

- **Evaluation metrics**

- Evaluation metrics are used to measure how good the predictions of the system are

- The most important evaluation metrics for classification include **Accuracy, Precision, Recall, F1 Score, Specificity, and Sensitivity**

Classification

- **Evaluation metrics**

Evaluation Matrix

Confusion Matrix

	Actual Positive (Apple)	**Actual Negative (Non-apple)**
Predicted Positive (Apple)	TP	FP
Predicted Negative (Non-apple)	FN	TN

TP: true positive
FP: false positive
FN: false negative
TN: true negative

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

NOTE: Accuracy is NOT always a good measure!

Classification

- **Evaluation metrics**

Confusion Matrix

Total fruits: 20

	Actual Positive (Apple)	**Actual Negative (Non-apple)**
Predicted Positive (Apple)	10 TP	5 FP
Predicted Negative (Non-apple)	2 FN	3 TN

TP: true positive
FN: false negative

FP: false positive
TN: true negative

Classification

- **Evaluation metrics**

TP: The number of positive samples that are predicted correctly as positive (**T**ruly classified as **P**ositive)

TN: The number of negative samples that are predicted correctly as negative (**T**ruly classified as **N**egative)

FP: The number of negative samples that are predicted incorrectly as positive (**F**alsely classified as **P**ositive)

FN: The number of positive samples that are predicted incorrectly as negative (**F**alsely classified as **N**egative)

Classification

- **Evaluation metrics**

Limitation

Suppose

- Total number of fruits in the testing examples = 10,000
- Number of Non-apple = 9990
- Number of Apple = 10

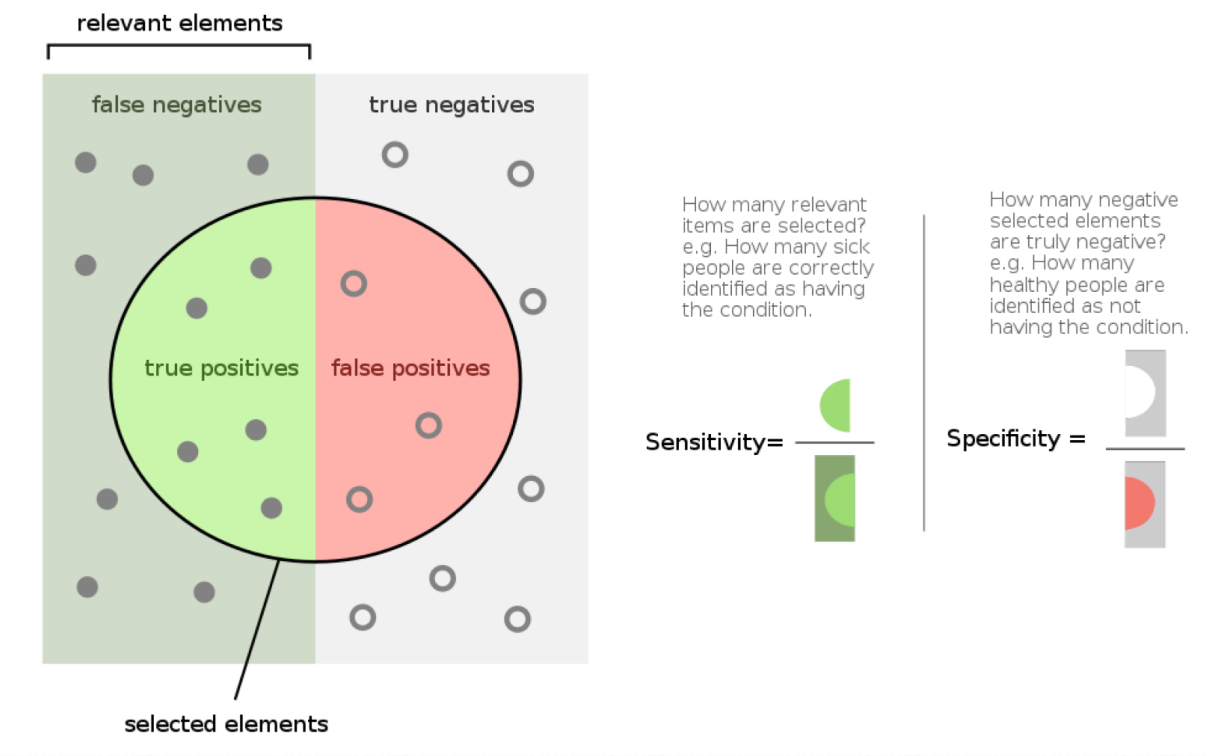
Can you classify Apple?

If model predicts everything to be of class non-apple, the accuracy is 9990/10000 = 99.9 %!!!

Here, accuracy is misleading because model cannot detect any Apple at all, still achieving high accuracy.

Classification

- **Evaluation metrics**



Sensitivity and specificity. (2024, January 24). In *Wikipedia*. https://en.wikipedia.org/wiki/Sensitivity_and_specificity

Classification

- **Evaluation metrics**

sensitivity, recall, hit rate, or true positive rate (TPR)

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

specificity, selectivity or true negative rate (TNR)

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{N}} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}$$

precision or positive predictive value (PPV)

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

Sensitivity and specificity. (2024, January 24). In *Wikipedia*.
https://en.wikipedia.org/wiki/Sensitivity_and_specificity

Classification

- **Evaluation metrics**

Evaluation Matrix

	Actual Positive (Apple)	Actual Negative (Non-apple)
Predicted Positive (Apple)	TP	FP
Predicted Negative (Non-apple)	FN	TN

$$Precision = \frac{TP}{TP + FP}$$

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

$$Recall = \frac{TP}{TP+FN}$$

Classification

- **Evaluation metrics**

NOTE 1: These metrics are defined for binary classification, but can be extended to multi-class classification

NOTE 2: It is optional to consider either of the classes as negative or positive. However, once you have defined a class as negative or positive, you should stick to your definition.

Classification

- **Evaluation metrics**

NOTE 1: These metrics are defined for binary classification, but can be extended to multi-class classification

NOTE 2: It is optional to consider either of the classes as negative or positive. However, once you have defined a class as negative or positive, you should stick to your definition.

Classification

- **Evaluation metrics - example**

Question: Assume we have 1000 fruits, and we want to classify them into two classes as follows: Apple and Non-apple. Assume we have 10 apples in the dataset, and the rest of the fruits are not apples. If the classifier classifies all fruits as Non-apple, calculate the evaluation metrics for this classifier.

Solution: We define the Non-apple class as positive and the Apple class as negative.
TP= 990 (How many Non-apple fruits are classified as Non-apple)
TN= 0 (How many Apple fruits are classified as Apple)
FP= 10 (How many Apple fruits are classified as Non-apple)
FN= 0 (How many Non-apple fruits are classified as Apple)

Classification

- **Evaluation metrics - example**

Accuracy = (TP+TN)/(TP+TN+FP+FN)=(990+0)/(990+0+10+0)=0.99 (99%)
Interpretation: The model is prediction 99% of cases correctly.

Sensitivity= TP/(TP+FN)= 990/(990+0)=1 (100%)
Interpretation: The model is predicting 100% of Non-apple fruits correctly

Specificity= TN/(TN+FP)= 0/(0+10)=0 (0%)
Interpretation: The model is predicting 0% of Apple fruits correctly

Precision= TP/(TP+FP)= 990/(990+10)=0.99 (99%)
Interpretation: 99% of predicted Non-apple fruits are correct

Classification

- **Evaluation – Cross validation**

- In classification, we use a portion of data for training the model. This portion is called **training data**
- We use the rest of the data for resting the model. This portion is called **testing data**.
- There is a challenge. Which part of data should be used for training the model and which part should be used for testing?

Classification

- **Evaluation – Cross validation**

- In the cross-validation (CV) method, we divide data into K folds. Then in each iteration, one of the folds is reserved for training the model and one portion is reserved for testing the model.
- We repeat this process until all folds have been used for testing at least once.
- This procedure is also called K-fold cross-validation.
- The accuracy of the model will be the average accuracy of all iterations.
- The best practice is to test the model on separate test data after the cross-validation

Classification

- **Evaluation – Cross validation**