USTH
VIETNAM FRANCE UNIVERSITY

Lecture 4

Model Evaluation

Dr. Le Huu Ton

Hanoi, 2017

Model Evaluation

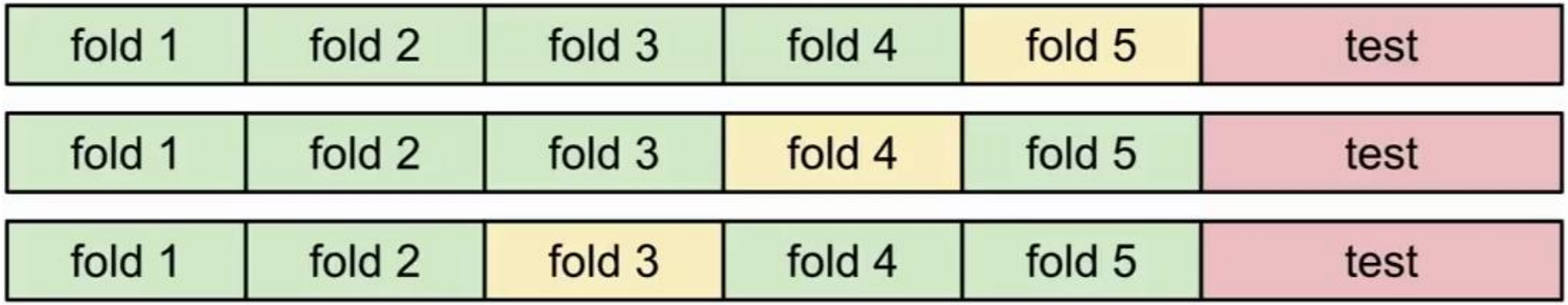
Idea #1: Choose hyperparameters that work best on the data

Idea #2: Split data into **train** and **test**, choose hyperparameters that work best on test data

Idea #3: Split data into **train**, **val**, and **test**; choose hyperparameters on val and evaluate on test

Model Evaluation

Idea #4: **Cross-Validation**: Split data into **folds**,
try each fold as validation and average the results



Useful for small datasets, but not used too frequently in deep learning

Model Evaluation

- **Training Dataset**: The sample of data used to fit the model.
- **Validation Dataset**: The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters. The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration.
- **Test Dataset**: The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset.

Evaluation of Regression Model

- Mean Absolute Error.
- Mean Squared Error.
- R Square (R^2)

Evaluation of Regression Model

Mean Absolute Error: is the sum of the absolute differences between predictions and actual values

$$MAE = \frac{1}{m}\sum_{i=1}^{m} |y^i - h(x^i)|$$

Evaluation of Regression Model

Mean Square Error: measures the average of the squares of the errors- the difference between the real value and the predicted one

$$MSE = \frac{1}{m}\sum_{i=1}^{m}\left(y^i - h(x^i)\right)^2$$

Evaluation of Regression Model

R Square: provides an indication of the goodness of fit of a set of predictions to the actual values. In statistical literature, this measure is called the coefficient of determination.

The total sum of squares (proportional to the variance of the data):

$$SS_{tot} = \sum_{i=1}^{m} \left(y^i - \bar{y}\right)^2$$

The sum of squares of residuals, also called the residual sum of squares

$$SS_{res} = \sum_{i=1}^{m} \left(y^i - h(x^i)\right)^2$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

Evaluation of Classification Model

- Confusion matrix
- Recall
- Precision
- Specificity
- Accuracy
- F-score
- Root mean squared error (RMSE)

Evaluation of Classification Model

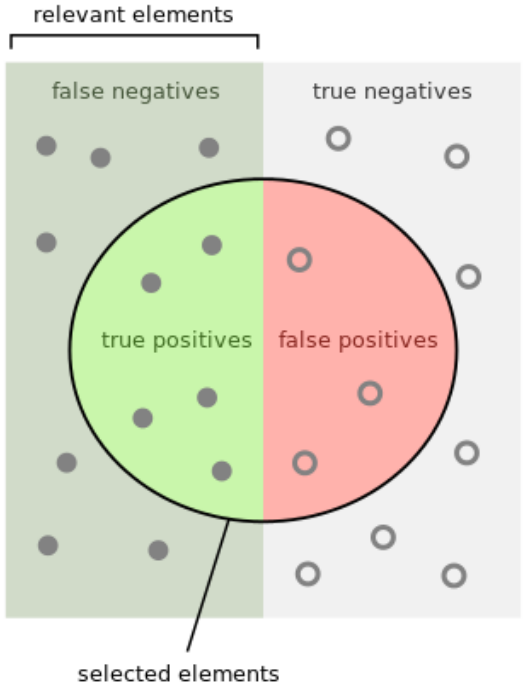
- Confusion matrix: a **confusion matrix**, also known as an error matrix, is a specific table layout that allows visualization of the performance of an algorithm

		Actual class		
		Cat	**Dog**	**Rabbit**
Predicted class	**Cat**	5	2	0
	Dog	3	3	2
	Rabbit	0	1	11

Evaluation of Classification Model

- Precision, Recall, Accuracy and Specificity

		Predicted Label	
		Positive	**Negative**
Known Label	**Positive**	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)



Measure	Formula	Intuitive Meaning
Precision	TP / (TP + FP)	The percentage of positive predictions that are correct.
Recall / Sensitivity	TP / (TP + FN)	The percentage of positive labeled instances that were predicted as positive.
Specificity	TN / (TN + FP)	The percentage of negative labeled instances that were predicted as negative.
Accuracy	(TP + TN) / (TP + TN + FP + FN)	The percentage of predictions that are correct.

Evaluation of Classification Model

- F-score:

$$F = 2 * \frac{precision * recall}{precision + recall}$$

References

http://openclassroom.stanford.edu/MainFolder/CoursePage.php?course=MachineLearning

https://en.wikipedia.org/wiki/Activation_function

https://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example/