# How can we know that a machine learning algorithm classifies and detects faults well?

Howard Cheung

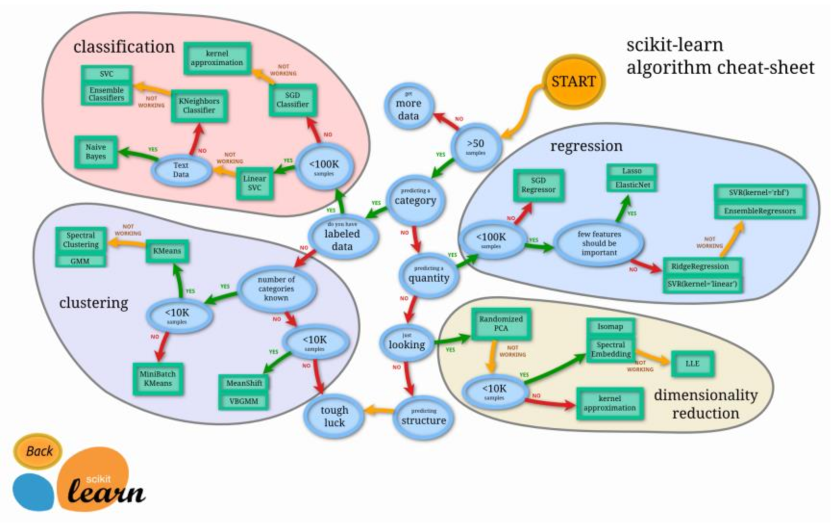
2018/10/25

# Agenda

- Classification: what can we do with it?
- What is a good classification algorithm?
  - Confusion matrix and Accuracy score
  - True and False Positive Rate
  - Receiver Operating Characteristic Curve
- Example on complex applications
- Conclusions

# Classification

- One major application of supervised learning methods in machine learning
- Work on categorized training data
  - Learn how categories vary with data
  - Categorize future data automatically



# Classification

- Applications
  - Prediction of future event outcome

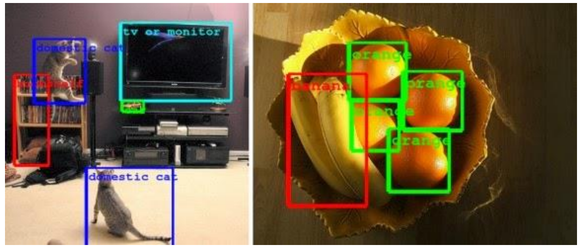
| Exam 1 grade | Quiz 1 grade | Course grade |
|---|---|---|
| 80 | 50 | Yes |
| 50 | 70 | No |

| Exam 1 grade | Quiz 1 grade | Course grade |
|---|---|---|
| 50 | 20 | ? |
| 60 | 30 | ? |

  - Fault detection and diagnostics

Normal?
Stator fault?
Rotor fault?

  - Image recognition



# Classification

- Example algorithms
  - Logistic Regression Classification
  - Support Vector Classification
  - Naïve Bayes Classification

- Their end products are similar
  - Predicted categories
  - Hence similar evaluation method

# Good classification algorithm

- Predict the classes correctly

- Consider the confusion matrix of the training data

| | Class 1 data | Class 2 data | Class 3 data |
|---|---|---|---|
| **Predicted to be class 1** | 50 | 30 | 10 |
| **Predicted to be class 2** | 10 | 60 | 5 |
| **Predicted to be class 3** | 20 | 10 | 80 |

- More counts on the diagonal, more accurate a classification is

# But is that how it should be?

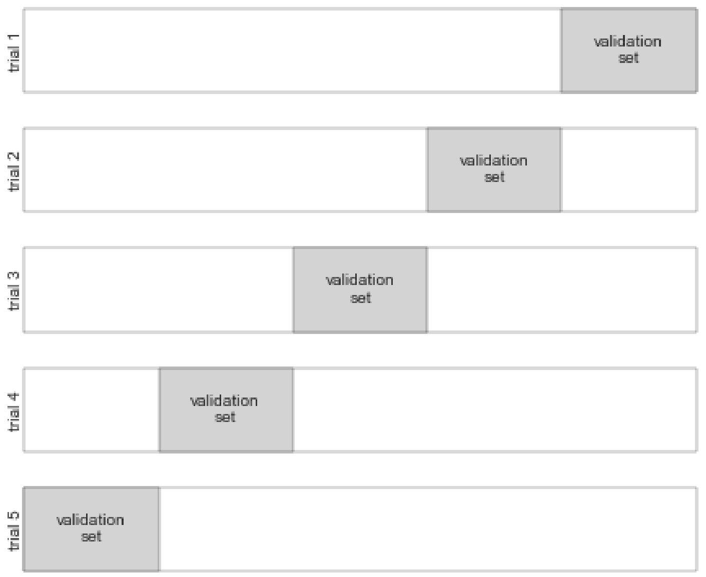
- There are quite a few problems in the previous analysis
  - Used all training data for evaluation
    - Not simulating what will happen to cases not used for training
  - Not application specific
    - Not all classes are equal
  - On a specific setup
    - What will happen if the user wants a set of stricter/ more relaxed rules?

# How can we address the issues?

- Cross-validation
- True and false positive rates
- Receiver-operating characteristic (ROC) curve

# Cross-validation

- Separate training datasets into multiple subsets
- Get accuracy scores of all trials with their own validation sets
- Examine the performance of the algorithms in cases outside their training data



VanderPlas (2016)

# True and false positive rates

- Not all classes are equal!

- Example: Fault Detection and Diagnostics
  - Sometimes classification algorithms are only used to evaluate if a minor issue occurs
    - Send someone to repair the issue if a fault happens

  - But sending someone to fix some non-existing issues are costly
    - Worse than sending no one at all!

  - You want to check the accuracy of the classes differently

# True and false positive rates

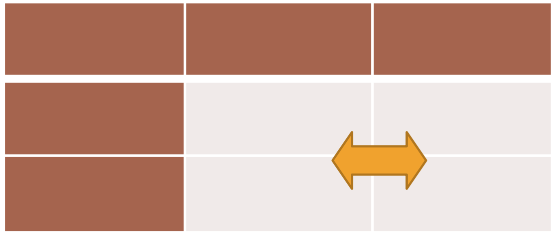
- Consider a confusion matrix

| | Predicted to be class 1 | Predicted to be other classes |
|---|---|---|
| Really belongs to class 1 | True positive | False negative |
| It's not in class 1 | False positive | True negative |

- If a class 1 prediction requires costly actions, you will need a algorithm
  - Very low false positive rate $= \frac{\text{Number of false positive cases}}{\text{Number of non}-\text{class 1 cases}}$
  - Moderately high true positive rate $= \frac{\text{Number of true positive cases}}{\text{Number of actual class 1 cases}}$

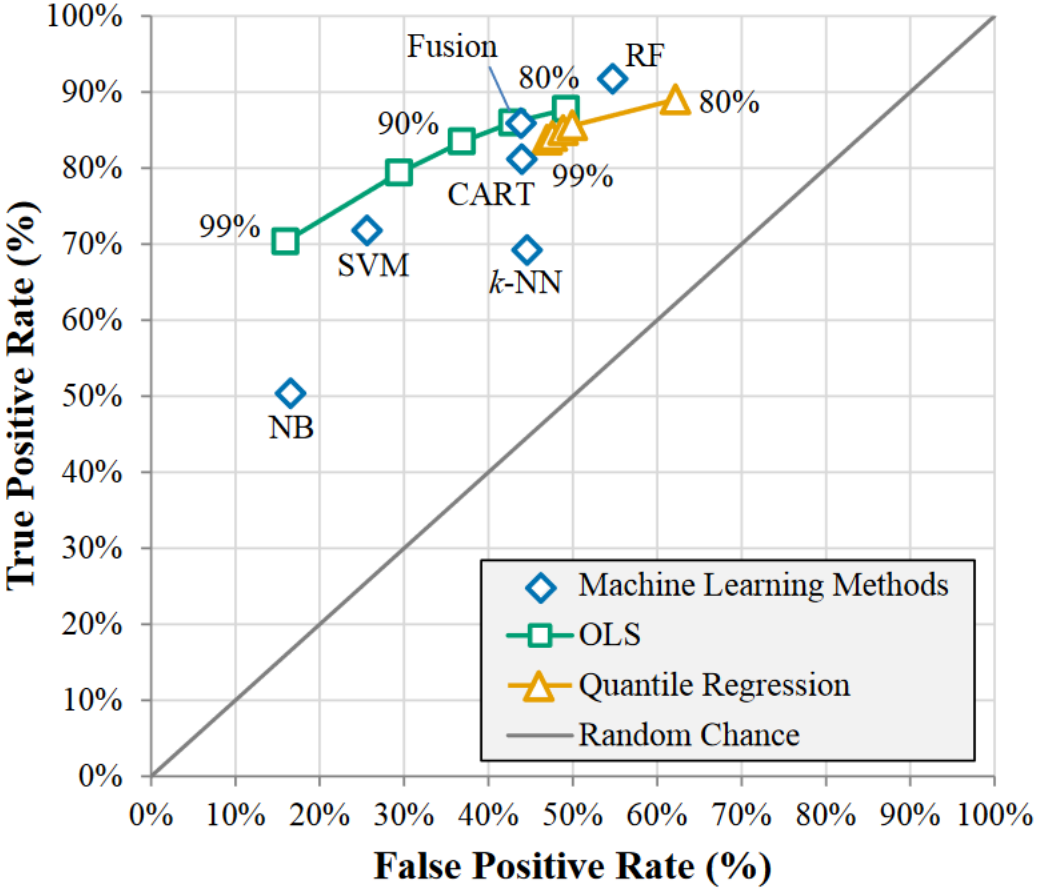
# Receiver operating characteristic curves

- There are thresholds to adjust the criteria between the boundaries of the classes
  - Regularization strength to avoid overfitting at the cost of accuracy
  - Confidence level



- Change of threshold changes the performance of the algorithm
- Range of performance metrics instead of just one metric

# ROC curve




Frank et al. (2016)

# Study of a problematic case

- There are some algorithms trying to claim very different classification rules to game the system
- Claim: “Our algorithm is not so good at the identification of Issues Y and Z. If Issue Y exists, it may raise either Issue Y or Z alarm. But not vice versa.”
- Will that work better?

# Let's look at the original confusion matrices without the claim

- Since we have two diagnoses for two issues, we have two matrices

- Matrix for issue Y prediction
  - Where are the true positives for issue Y classification?
  - 4 True positives and 4 False positives

| | | Predicted to have both Y and Z | Predicted to have Y but not Z | Predicted to have Z but not Y | Predicted not to have Y and Z |
|---|---|---|---|---|---|
| With issue Y | With issue Z | True positive (TP) | | False negative (FN) | |
| | Without issue Z | | | | |
| Without issue Y | With issue Z | False positive (FP) | | True negative (TN) | |
| | Without issue Z | | | | |

# Let's look at the original confusion matrices without the claim

- Matrix for issue Z prediction
  - Where are the true positives for issue Z classification?
  - 4 TPs and 4FPs

| | | Predicted to have both Y and Z | Predicted to have Y but not Z | Predicted to have Z but not Y | Predicted not to have Y and Z |
|---|---|---|---|---|---|
| With issue Y | With issue Z | TP | FN | TP | FN |
| | Without issue Z | FP | TN | FP | TN |
| Without issue Y | With issue Z | TP | FN | TP | FN |
| | Without issue Z | FP | TN | FP | TN |

# Let’s see what will happen with the claim to the issue Z classification

- Matrix for issue Z prediction
  - WP stands for weak positive which means that the classification is partially correct

| | | Predicted to have both Y and Z | Predicted to have Y but not Z | Predicted to have Z but not Y | Predicted not to have Y and Z |
|---|---|---|---|---|---|
| With issue Y | With issue Z | TP | FN | WP | FN |
| | Without issue Z | FP | TN | WP | FN |
| Without issue Y | With issue Z | TP | FN | TP | FN |
| | Without issue Z | FP | TN | FP | TN |

# Why?

- If there are two issues but the algorithm is giving me one alarm, I am going to check for one issue only
  - One partially correct answer (WP)
- If the algorithm tells me that an issue Z exists but not issue Y, I am going to check issue Z first. If it doesn't exist, I go to check issue Y. <u>This is not automatic!</u>
  - Another partially correct answer (WP)
- You should give an issue Z prediction even if only issue Y exists
  - One TN turns into a FN

# So what happened?

- Convert
  - One True Positive scenario
  - One False Positive scenario
  - One True Negative scenario
- To
  - Two Weak Positive scenarios
  - One False Negative scenario
- Reduce
  - True Positive Rate, False Positive Rate and Accuracy Score
  - Only good if your false positive rate is too high

# Summary

- Classification algorithm evaluation methods
  - General
    - Accuracy score and confusion matrix
  - Specific applications
    - True and false positive rates
  - Tuneable parameters
    - ROC curve

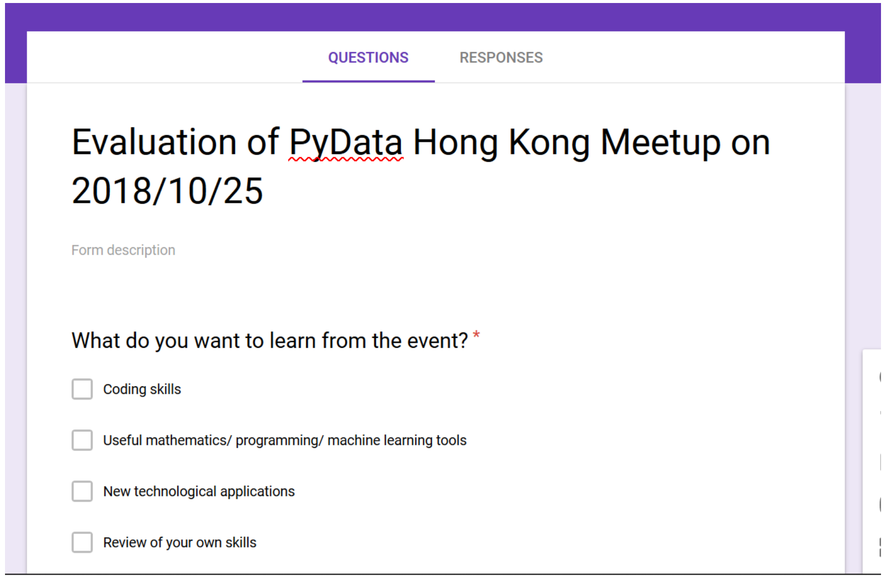
- Confusion matrix construction to evaluate special claims

# References


- Cheung, H., Braun, J. E., & Langner, M. R. (2016). *Methodology to Assess No Touch Audit Software Using Simulated Building Utility Data* (No. NREL/SR-5500-66001). Golden, CO: National Renewable Energy Laboratory. Retrieved from http://www.nrel.gov/docs/fy17osti/66001.pdf
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010
- Frank, S., Heaney, M., Jin, X., Robertson, J., Cheung, H., Elmore, R., & Henze, G. (2016). Hybrid Model-Based and Data-Driven Fault Detection and Diagnostics for Commercial Buildings. In 2016 ACEEE Summer Study on Energy Efficiency in Buildings. Retrieved from http://www.nrel.gov/docs/fy16osti/65924.pdf
- Kaggle. Titanic: Machine Learning from Disaster. Retrieved October 23, 2018, from https://www.kaggle.com/c/titanic
- VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data (1 edition). Sebastopol, CA: O'Reilly Media. Retrieved from https://jakevdp.github.io/PythonDataScienceHandbook/index.html

# Evaluation

- https://bit.ly/2SbvtiX


QUESTIONS RESPONSES

Evaluation of PyData Hong Kong Meetup on 2018/10/25

Form description

What do you want to learn from the event? *

- [ ] Coding skills
- [ ] Useful mathematics/ programming/ machine learning tools
- [ ] New technological applications
- [ ] Review of your own skills