CSCE 581: Introduction to Trusted AI

Lectures 7 and 8: Supervised ML, Project, Trust

PROF. BIPLAV SRIVASTAVA, AI INSTITUTE

4TH AND 6TH FEB, 2025

Carolinian Creed: "I will practice personal and academic integrity."

Credits: **Copyrights of all material reused acknowledged**

Organization of Lectures 7, 8

- Introduction Section
  - Recap from Week 3 (Lectures 5 and 6)
  - Announcements and News
- Main Section
  - L7: Supervised ML (contd)
  - L8: Project Discussion
  - Trust issues in (Supervised) ML
- Concluding Section
  - About next week – Lectures 9, 10
  - Ask me anything

Introduction Section

Recap from Week 3 (Lectures 5, 6)

- We looked at
  - Data and characteristics
  - Data organization, ontologies
  - ML background
- Project discussion

AI News

- DeepSeek R1 – a POV from our own analysis
(https://drive.google.com/file/d/1gKKM0sEcp5u6pA05jETZSCQZNJSN7SJt/view?usp=sharing)

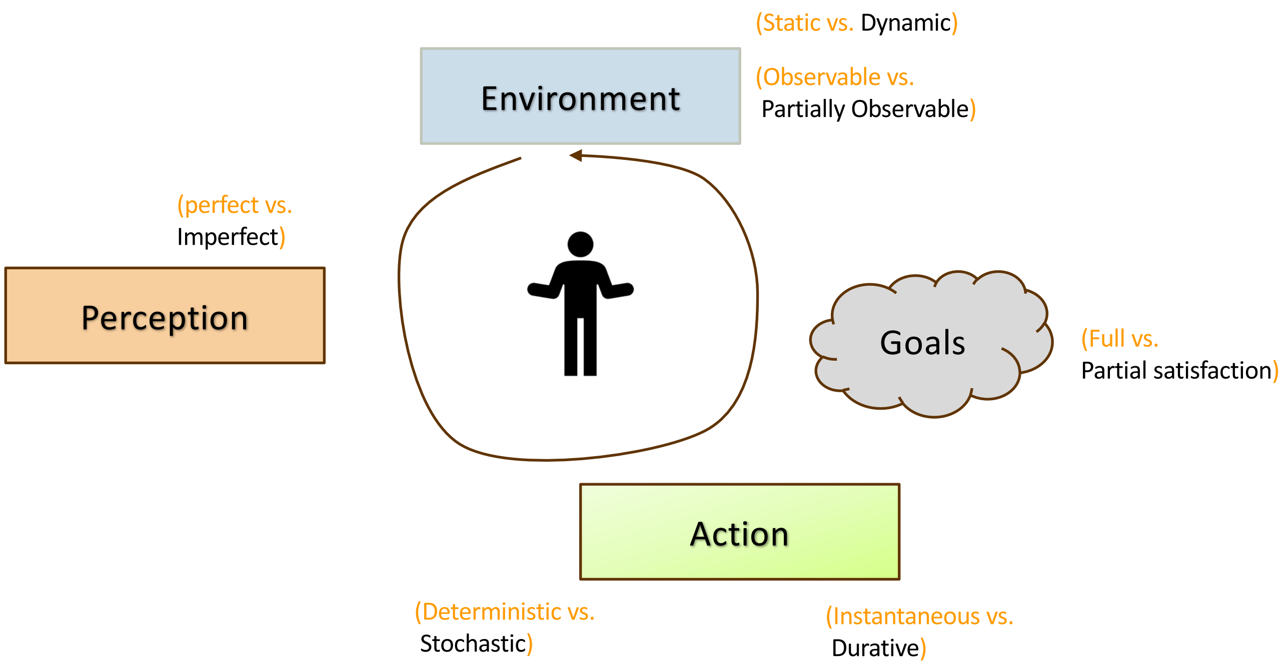
Announcement: Change to
Student Assessment

A = [920-1000]
B+ = [870-919]
B = [820-869]
C+ = [770-819]
C = [720-769]
D+ = [670-719]
D = [600-669]
F = [0-599]

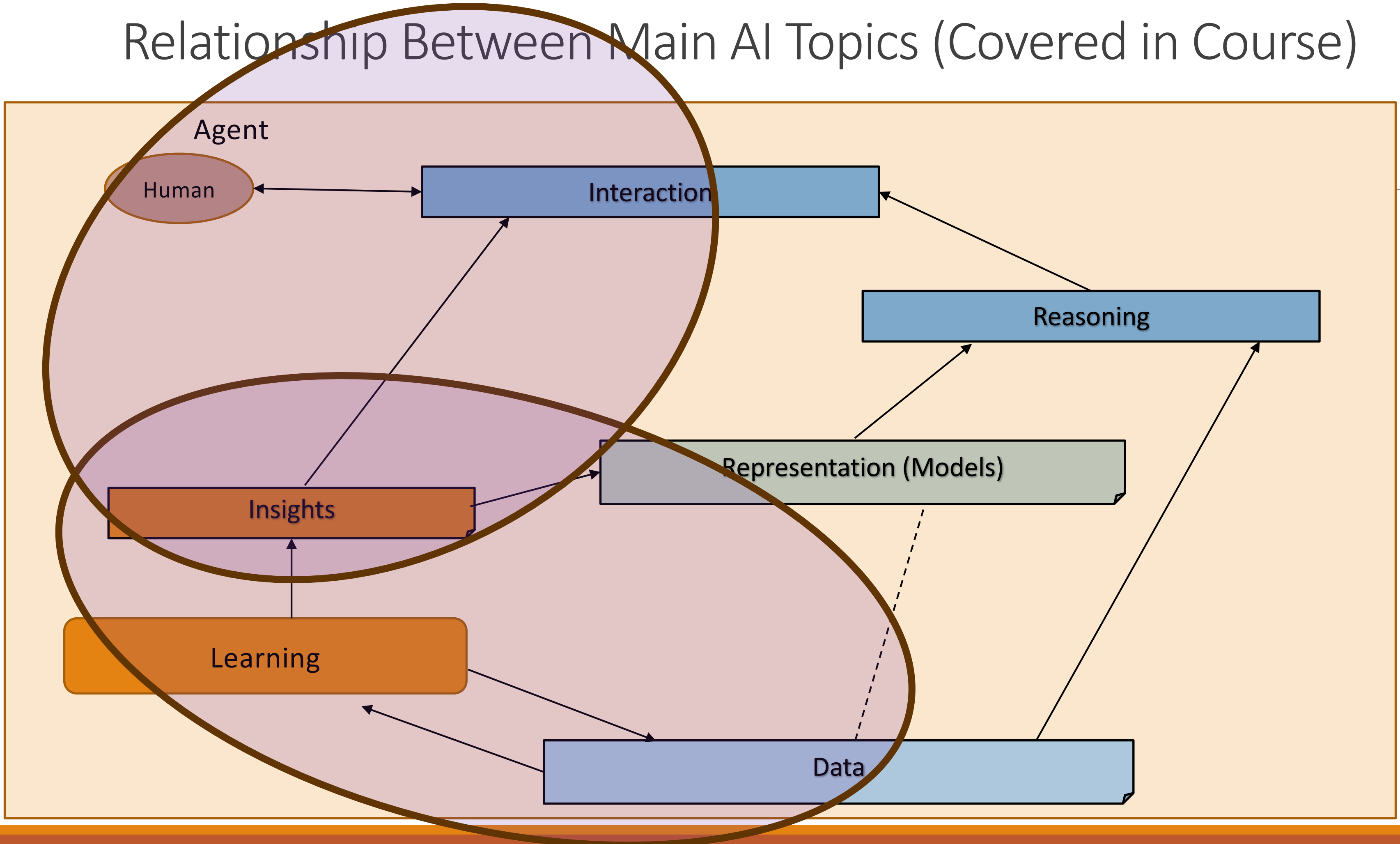
Tests	Undergrad	Grad
Course Project – report, in-class presentation	600	600
Quiz – 2 quizzes	200	200
Final Exam	200	100
Additional Final Exam – Paper summary, in-class presentation		100
Total	1000 points	1000 points

Change: 4 quizzes to 2; no best of 3

Intelligent Agent Model



Relationship Between Main AI Topics (Covered in Course)



High Level Semester Plan (Adapted, Approximate)

CSCE 581 –

- Week 1: Introduction
- Week 2: Background: AI - Common Methods
- Week 3: The Trust Problem
- Week 4: Machine Learning (Structured data) - Classification
- Week 5: Machine Learning (Structured data) - Classification – Trust Issues
- Week 6: Machine Learning (Structured data) – Classification – Mitigation Methods
- Week 7: Machine Learning (Structured data) – Classification – Explanation Methods
- Week 8: Machine Learning (Text data, **vision**) – Classification, **Large Language Models**
- Week 9: Machine Learning (Text data) - Classification – Trust Issues, LLMs
- Week 10: Machine Learning (Text data) – Classification – Mitigation Methods
- Week 11: Machine Learning (Text data) – Classification – Explanation Methods
- Week 12: Emerging Standards and Laws, **Real world applications**
- Week 13: Project presentations
- Week 14: Project presentations, Conclusion

AI/ ML topics and with a focus on fairness, explanation, Data privacy, reliability

Main Section

Machine Learning

Credit: Retrieved from internet

Machine Learning – Insights from Data

- Descriptive analysis
  - Describe a past phenomenon
  - **Methods**: classification (feedback from label), clustering, dimensionality reduction, anomaly detection, neural methods, reinforcement learning (feedback from hint/ reward)
- Predictive analysis
  - Predict about a new situation
  - **Methods**: time-series, neural networks
- Prescriptive analysis
  - What an agent should do
  - **Methods**: simulation, reinforcement learning, reasoning
- New areas
  - Counterfactual analysis
  - Causal Inferencing
  - Scenario planning

Concepts

- **Input data**: data available
  - **Training data**: used for training a learning algorithm and get a model
    - [Optional] **Validation data**: used to tune parameters
  - **Test data**: used to test a learning model

- **Classification problem**
  - Separating data into classes (also called labels, categorical types)
  - One of the attributes is the class label we are trying to learn
  - Class label is the **supervision**
- **Clustering problem**
  - We are trying to learn grouping of data
  - There is no attribute indicating membership in the groups (hence, **unsupervised**)
- **Prediction problem**
  - Learning value of a continuous variable

Reference: https://machinelearningmastery.com/difference-test-validation-datasets/
https://www2.seas.gwu.edu/~bell/csci243/lectures/classification.pdf

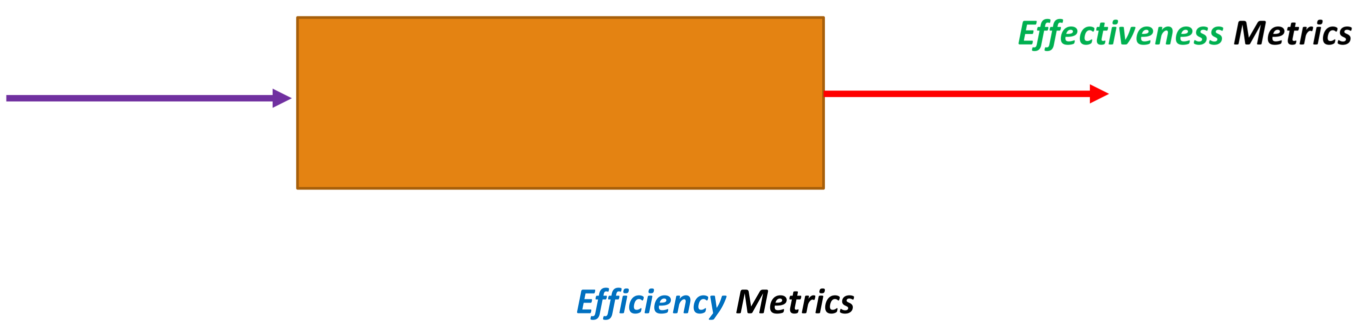
Sample Learning Task

- COVID-19 data

Notebook: https://github.com/biplav-s/course-d2d-ai/blob/main/sample-code/l6-l7-l8-supervised-ml/Supervised-Regression-Classification.ipynb

Metric Types

- **Effectiveness**: what the **user** of a system sees, primarily cares about
- **Efficiency**: what the **executor** in a system sees, primarily cares about



Metrics: Accuracy, Precision, Recall

	Predicted class		
Actual Class		Class = Yes	Class = No
	Class = Yes	True Positive	False Negative
	Class = No	False Positive	True Negative

Accuracy =
(TP+TN)/
(TP+FP+FN+TN)

Precision =
(TP)/
(TP+FP)

Recall =
(TP)/
(TP+FN)

F1 Score: *Harmonic Mean*

1/F1 = 1/Precision + 1/Recall

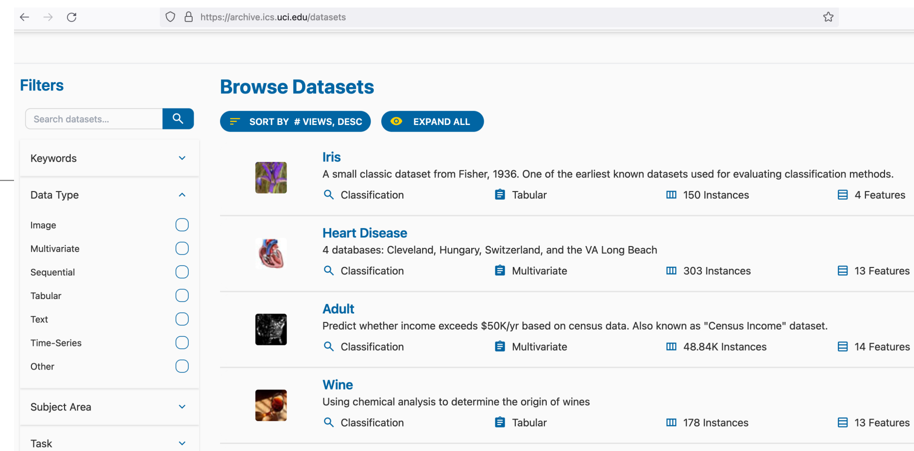
F1 = 2*(Recall * Precision) /
(Recall + Precision)

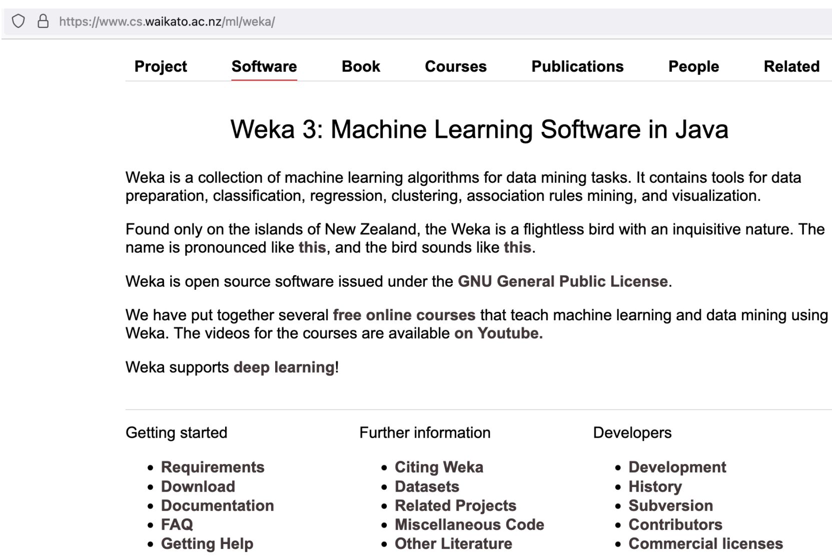
Example: Detecting Spam in Email

- **Effectiveness**: what the **user** of a system sees, primarily cares about
  - *How many spams identified?*
  - *How many spams missed?*
- **Efficiency**: what the **executor** in a system sees, primarily cares about
  - *How fast were spams detected?*
  - *How much memory was used per million emails processed ?*

Reference and Demo

- Data: UCI Datasets
  - https://archive.ics.uci.edu/datasets
  - Browse or search



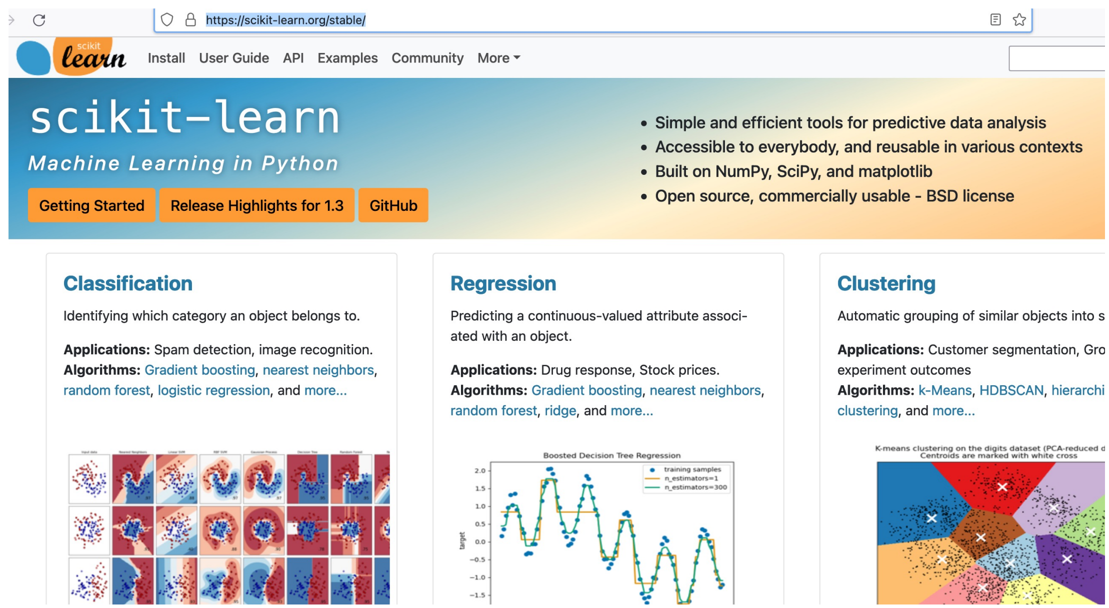


- Tools:
  - Weka - https://www.cs.waikato.ac.nz/ml/weka/
  - Download tool and dataset

- Libraries
  - Scikit - https://scikit-learn.org/stable/

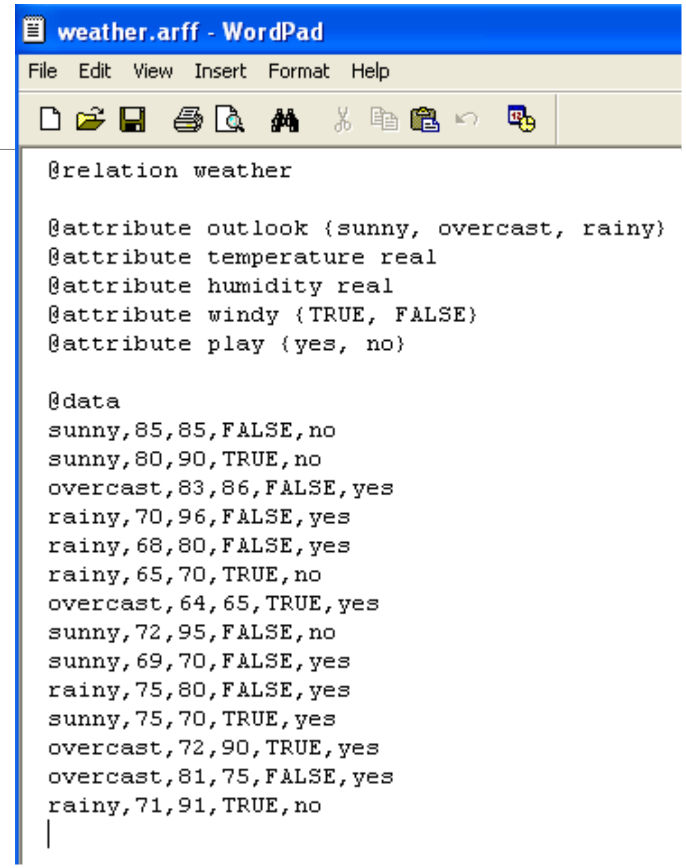
Reference and Demo

- Data: UCI Datasets
  - https://archive.ics.uci.edu/datasets
  - Browse or search
- Tools:
  - Weka - https://www.cs.waikato.ac.nz/ml/weka/
  - Download tool and dataset
- Libraries
  - Scikit - https://scikit-learn.org/stable/



ARFF Data Format

- Attribute-Relation File Format
- Header – describing the attribute types
- Data – (instances, examples) comma-separated list




```
@relation weather

@attribute outlook {sunny, overcast, rainy}
@attribute temperature real
@attribute humidity real
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@data
sunny,85,85,FALSE,no
sunny,80,90,TRUE,no
overcast,83,86,FALSE,yes
rainy,70,96,FALSE,yes
rainy,68,80,FALSE,yes
rainy,65,70,TRUE,no
overcast,64,65,TRUE,yes
sunny,72,95,FALSE,no
sunny,69,70,FALSE,yes
rainy,75,80,FALSE,yes
sunny,75,70,TRUE,yes
overcast,72,90,TRUE,yes
overcast,81,75,FALSE,yes
rainy,71,91,TRUE,no
```

Slide Courtesy: http://www.cs.iastate.edu/~cs573x/bbsilab.html

Water Data – Water Atlas

- Data download: https://dev.chnep.wateratlas.usf.edu/data-download/beta/

- Local cache: https://github.com/biplav-s/course-tai/tree/main/sample-code/common-data/water

Exercise: German Credit

Datasets

- UCI Dataset:
  - Weka: https://www.ics.uci.edu/~mlearn/MLRepository.html (e.g., download: https://prdownloads.sourceforge.net/weka/uci-20070111.tar.gz)
  - Check in UCI – variants:
  - https://archive.ics.uci.edu/dataset/573/south+german+credit+update

- Weka
  - Direct link: https://github.com/Waikato/weka-3.8/blob/master/wekadocs/data/credit-g.arff
  - As part of development packages
    - like DataHub, https://datahub.io/machine-learning/credit-g#python

German Credit Data

https://archive.ics.uci.edu/ml/datasets/Statlog+%28German+Credit+Data%29

- Dataset that classifies people's credit risk based on their individual attributes such as Age, Income, Gender, etc.
  - 1000 rows of data, each with 20 attributes to check bias against
- Each entry represents an individual who takes credit from a bank
- Each entry is classified as *Good* or *Bad* credit risk based on their profile

1. Credit amount (numerical);
2. Credit duration (numerical);
3. Credit purpose (categorical);
4. Status of existing checking account(categorical);
5. Status of savings accounts and bonds (categorical);
6. Number of existing credits (numerical);
7. Credit history(categorical);
8. Installment plans (categorical);
9. Installment rate(numerical);
10. Property (categorical);
11. Residence (categorical);
12. Period of present residency (numerical);
13. Telephone (binary);
14. Employment (categorical);
15. Employment length (categorical);
16. Personal status and gender (categorical); 1
17. Age (numerical);
18. Foreign worker (binary);
19. Dependents (numerical);
20. Other debtors (categorical);
21. Credit score (binary)

Example Instance:
A11 6 A34 A43 1169 A65 A75 4 A93 A101 4 A121 67 A143 A152 2 A173 1 A192 A201 1

Example record: Alice is requesting a loan amount of 1567 DM for a duration of 12 months for the purpose of purchasing a television, with a positive checking account balance that is smaller than 200 DM, having less than 100 DM in savings account, and having one existing credit at this bank. She duly paid existing credits at the bank till now and has no other installment plan. She possesses a car and owns a house, has been living at the present residence for one year and has a registered telephone. She is a skilled employee, working in the present employment for past four years. She is a 22-year-old married female and is a German citizen. She has one dependent and no guarantors. The recorded outcome for Alice (attribute #21) is a good credit score.

Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml], Irvine, CA: University of California, School of Information and Computer Science

VERMA, S., AND RUBIN, J. 2018. FAIRNESS DEFINITIONS EXPLAINED. IN PROCEEDINGS OF THE INTERNATIONAL WORKSHOP ON SOFTWARE FAIRNESS, FAIRWARE '18, 1–7. NEW YORK, NY, USA: ASSOCIATION FOR COMPUTING MACHINERY, HTTPS://WWW.ECE.UBC.CA/~MJULIA/PUBLICATIONS/FAIRNESS_DEFINITIONS_EXPLAINED_2018.PDF

German Credit Data

https://archive.ics.uci.edu/ml/datasets/Statlog+%28German+Credit+Data%29

- Dataset that classifies people's credit risk based on their individual attributes such as Age, Income, Gender, etc.
  - 1000 rows of data, each with 20 attributes to check bias against
- Each entry represents an individual who takes credit from a bank
- Each entry is classified as *Good* or *Bad* credit risk based on their profile
    - It is **worse** to class a **customer as good when they are bad**, than it is to class a **customer as bad when they are good**.

1. Credit amount (numerical);
2. Credit duration (numerical);
3. Credit purpose (categorical);
4. Status of existing checking account(categorical);
5. Status of savings accounts and bonds (categorical);
6. Number of existing credits (numerical);
7. Credit history(categorical);
8. Installment plans (categorical);
9. Installment rate(numerical);
10. Property (categorical);
11. Residence (categorical);
12. Period of present residency (numerical);
13. Telephone (binary);
14. Employment (categorical);
15. Employment length (categorical);
16. Personal status and gender (categorical); 1
17. Age (numerical);
18. Foreign worker (binary);
19. Dependents (numerical);
20. Other debtors (categorical);
21. Credit score (binary)

Example record: Alice is requesting a loan amount of 1567 DM for a duration of 12 months for the purpose of purchasing a television, with a positive checking account balance that is smaller than 200 DM, having less than 100 DM in savings account, and having one existing credit at this bank. She duly paid existing credits at the bank till now and has no other installment plan. She possesses a car and owns a house, has been living at the present residence for one year and has a registered telephone. She is a skilled employee, working in the present employment for past four years. She is a 22-year-old married female and is a German citizen. She has one dependent and no guarantors. The recorded outcome for Alice (attribute #21) is a good credit score.

Review detailed data exploration at:
https://www.kaggle.com/sanyalush/predicting-credit-risk

Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml], Irvine, CA: University of California, School of Information and Computer Science

VERMA, S., AND RUBIN, J. 2018. FAIRNESS DEFINITIONS EXPLAINED. IN PROCEEDINGS OF THE INTERNATIONAL WORKSHOP ON SOFTWARE FAIRNESS, FAIRWARE '18, 1–7. NEW YORK, NY, USA: ASSOCIATION FOR COMPUTING MACHINERY, HTTPS://WWW.ECE.UBC.CA/~MJULIA/PUBLICATIONS/FAIRNESS_DEFINITIONS_EXPLAINED_2018.PDF

Classification Methods

Linear Methods

Assumption: target value (y) is expected to be a linear combination of the features (Xj).

Function estimate (linear)
W: weight, b: bias

$$f(X_j) = X_j W + b$$

Error Term (mean squared error)

$$MSE = \frac{1}{n} \sum_{j=1}^{n} \left[f(X_{j\cdot}) - y_j \right]^2$$

Many variants depending on the nature of error being minimized: overfitting (Ridge), number of non-zero coefficients (Lasso), …

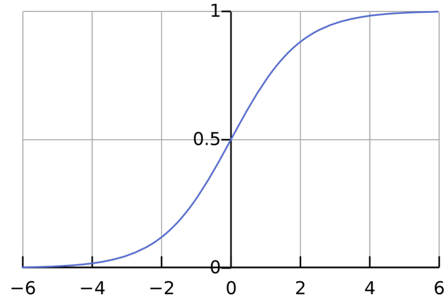
- Reference: https://scikit-learn.org/stable/modules/linear_model.html

Relationship Between Linear Regression and Classification

- Model type
  - Regression – Linear Regression:
    - predicting a continuous valued attribute assuming linear combination of feature vectors
  - Classification – Logistic Regression
    - Classifying a categorical attribute assuming linear combination of feature vectors
- Logit function

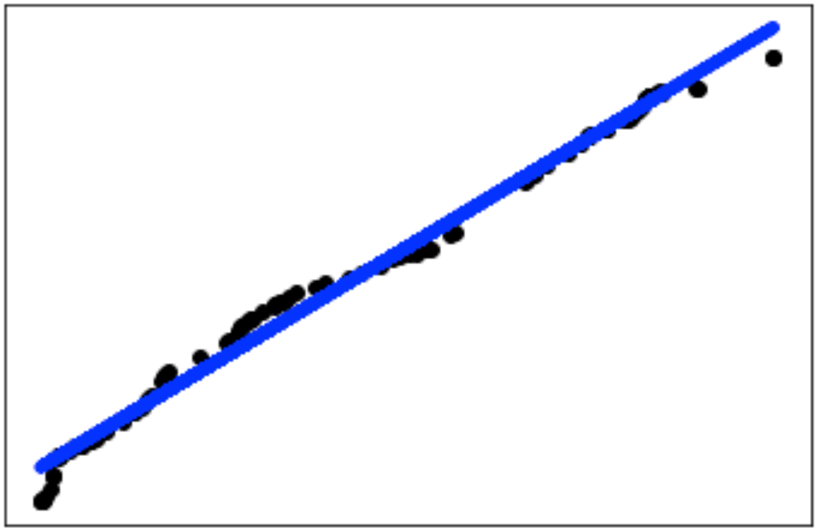
Example*: t is a linear function of a single explanatory variable x*

$$p(x) = \sigma(t) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$



Source: https://en.wikipedia.org/wiki/Logistic_regression

Linear Regression



Notebook: https://github.com/biplav-s/course-tai/blob/main/sample-code/l4-l5-supervised-ml/Supervised-Regression-Classification.ipynb

Reference and Demo

- Data: UCI Datasets - https://archive.ics.uci.edu/ml/datasets.php
- Tools:
  - Weka - https://www.cs.waikato.ac.nz/ml/weka/
    - ARFF format – Used by WEKA

Decision Tree

Problem: Classify Weather Data

Outlook	Temperature	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
…	…	…	…	…

Class Label

Input

Output (Informal)

```
If outlook = sunny and humidity = high then play = no
If outlook = rainy and windy = true then play = no
If outlook = overcast then play = yes
If humidity = normal then play = yes
If none of the above then play = yes
```

Slide Adapted From/ Courtesy: Data Mining, Practical Machine Learning Tools and Techniques, Slides for Chapter 6, Trees and Rules of Data Mining by I. H. Witten, E. Frank, M. A. Hall and C. J. Pal

Which Variable to Learn to Create Rules On?

- What do we want?
  - Compact model (e.g., set of rules)
  - High accuracy / low error
- Find the most discriminating variable
  - But how do we measure this

Outlook	Temperature	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
...	...	...	...	...

- Corner cases
  - If all the samples are the same, the decision tree is a ?
    - Leaf node with the only class
  - If there are no attributes in the dataset, the decision tree is?
    - A node with most common class

Expected Information/ Entropy

- Concept: Expected Information
  - Let
    - Class label has **m** distinct values (i.e., m distinct classes)
    - s_i be the number of samples of S of Class C_i (i = 1 ..m)
  - $I(s_1, s_2, .., s_m) = - \Sigma_{i=1 \text{ to } m}\, p_i \log_2 (p_i)$
    - Where P_i is the probability a sample belongs to class Ci; estimated by(s_i /s)

Outlook	Temperature	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
...	...	...	...	...

- Entropy / Expected Information after partitioning on Attribute A which has **v** distinct values
  - $E(A) = \Sigma_{j=1 \text{ to } v}\ (s_{1j} + \ldots + s_{mj}) / S \quad * (I(s_{1j}, s_{2j}, .., s_{mj})$
  - s_{ij} be the number of samples in S_j of Class C_i (i = 1 ..m)
  - Smaller the entropy, the greater the purity of the subset partitions

Illustrative Example

- Entropy before: 5 blue, 5 green nodes

$$E_{before} = -(0.5 \log_2 0.5 + 0.5 \log_2 0.5)$$
$$= \boxed{1}$$

- Entropy at split
  - A: left: 4 blue, right: 1 blue, 5 green

$$E_{left} = \boxed{0}$$

$$E_{right} = -(\frac{1}{6} \log_2(\frac{1}{6}) + \frac{5}{6} \log_2(\frac{5}{6}))$$
$$= \boxed{0.65}$$

  - Weigh entropy by size of sample in both nodes

$$E_{split} = 0.4 * 0 + 0.6 * 0.65$$
$$= \boxed{0.39}$$

Information gain:

$$\text{Gain} = 1 - 0.39 = \boxed{0.61}$$

Source: https://victorzhou.com/blog/information-gain/

Information Gain

Outlook	Temperature	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
...	...	...	...	...

- Entropy / Expected Information after partitioning on Attribute A which has v distinct values
  - $E(A) = \Sigma_{j=1 \text{ to } v} \ (s_{1j} + \ldots + s_{mj}) / S \ \ * (I(s_{1j}, s_{2j}, .., s_{mj})$
    - s_{ij} be the number of samples in S_j of Class C_i (i = 1 ..m)

- After partition, S_j
  - $I(s_{1j}, s_{2j}, .., s_{mj}) = - \Sigma_{i=1 \text{ to } m} \, p_{ij} \log_2 (p_{ij})$
  - Where p_{ij} is the probability a sample in S_j belongs to class C_i; estimated by$(s_{ij} / | s_j |)$

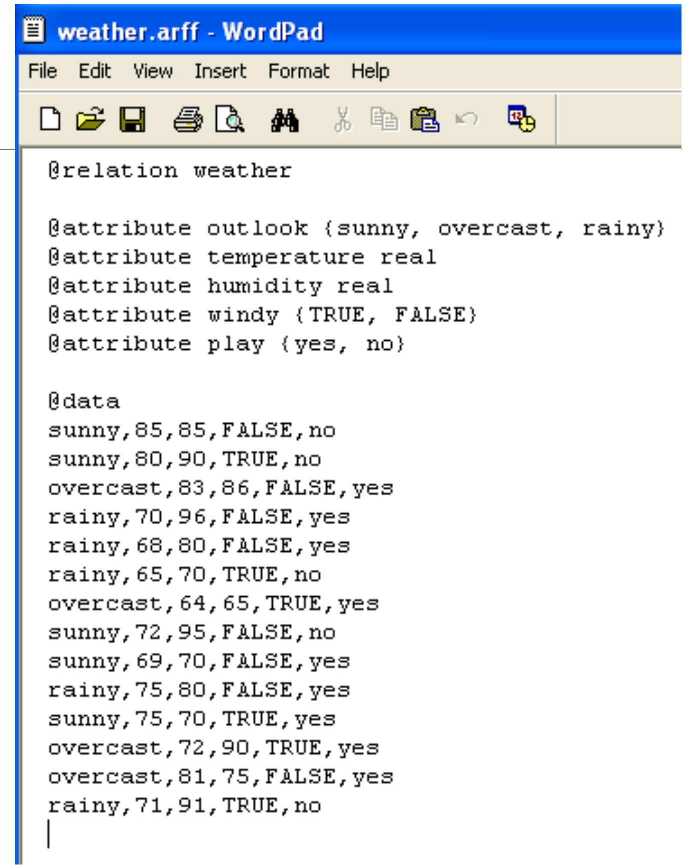
- Gain (A) = $I(s_1, s_2, .., s_m) - E(A)$
  - Is the expected reduction in entropy by knowing the value of Attribute A

- **Method**: Split on the attribute which leads to the highest information gain

Weka Exercise

ARFF Data Format

- Data is in ARFF in UCI dataset
- Or Convert
  - File system, CSV → ARFF format
  - Use [C45Loader]() and [CSVLoader]() to convert

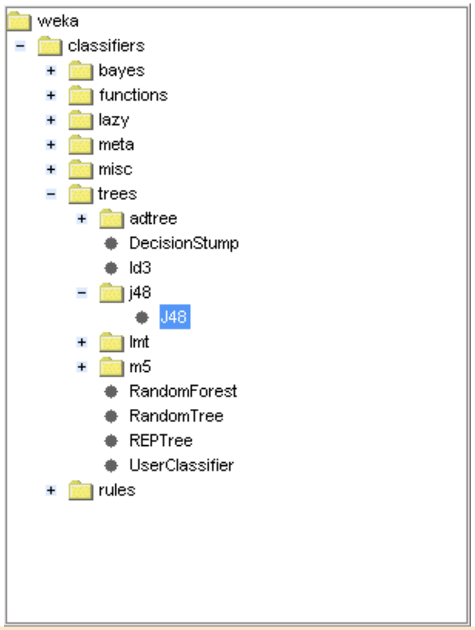


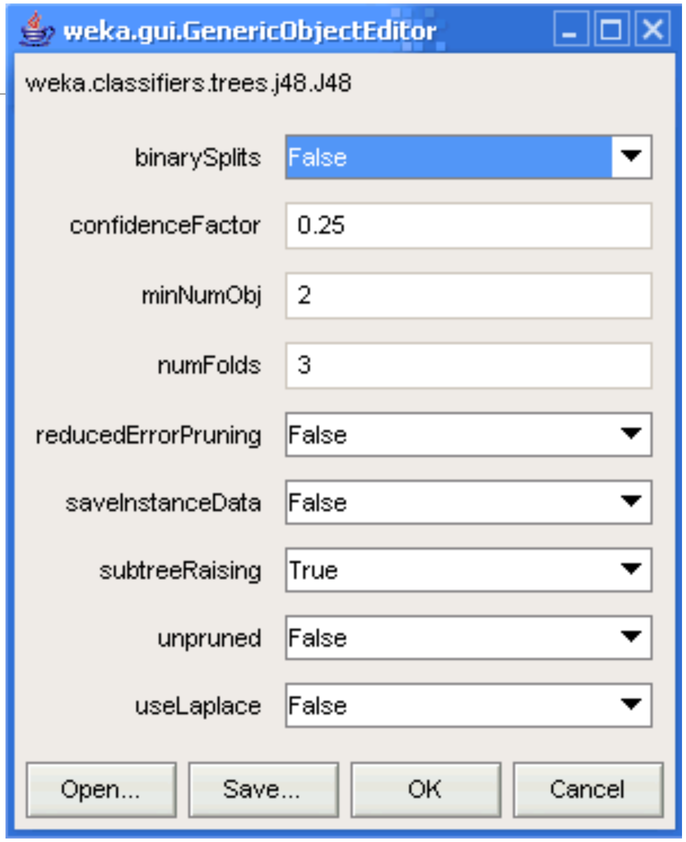

```
@relation weather

@attribute outlook {sunny, overcast, rainy}
@attribute temperature real
@attribute humidity real
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@data
sunny,85,85,FALSE,no
sunny,80,90,TRUE,no
overcast,83,86,FALSE,yes
rainy,70,96,FALSE,yes
rainy,68,80,FALSE,yes
rainy,65,70,TRUE,no
overcast,64,65,TRUE,yes
sunny,72,95,FALSE,no
sunny,69,70,FALSE,yes
rainy,75,80,FALSE,yes
sunny,75,70,TRUE,yes
overcast,72,90,TRUE,yes
overcast,81,75,FALSE,yes
rainy,71,91,TRUE,no
```

Slide Courtesy: http://www.cs.iastate.edu/~cs573x/bbsilab.html

Weka: weka.classifiers.trees.J48

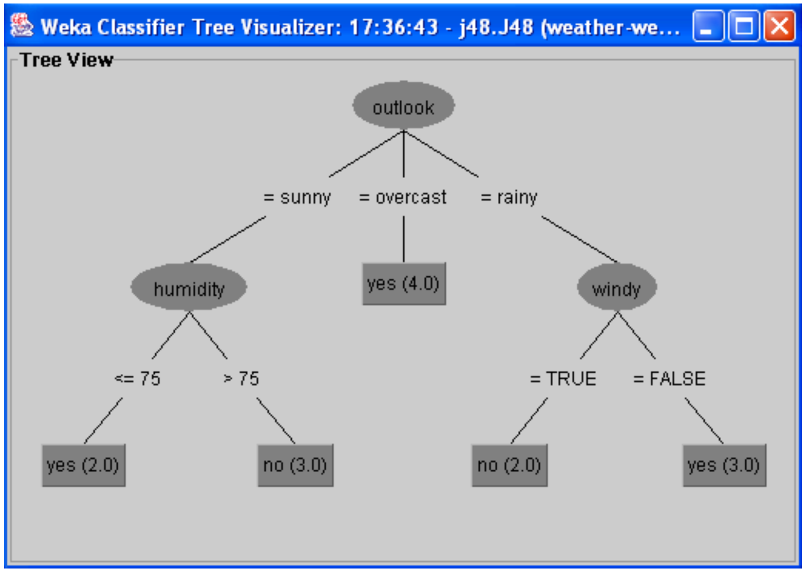
Class for generating an unpruned or a pruned C4.5 decision tree.





Slide Courtesy: http://www.cs.iastate.edu/~cs573x/bbsilab.html

Understanding Output



Slide Courtesy: http://www.cs.iastate.edu/~cs573x/bbsilab.html

Weka: Decision Tree Output

```
J48 pruned tree
------------------

outlook = sunny
|   humidity = high: no (3.0)
|   humidity = normal: yes (2.0)
outlook = overcast: yes (4.0)
outlook = rainy
|   windy = TRUE: no (2.0)
|   windy = FALSE: yes (3.0)

Number of Leaves  :    5

Size of the tree :       8
```

```
=== Summary ===

Correctly Classified Instances          7               50      %
Incorrectly Classified Instances        7               50      %
Kappa statistic                        -0.0426
Mean absolute error                     0.4167
Root mean squared error                 0.5984
Relative absolute error                87.5    %
Root relative squared error           121.2987 %
Total Number of Instances              14

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.556     0.6        0.625     0.556     0.588      0.633    yes
                 0.4       0.444      0.333     0.4       0.364      0.633    no
Weighted Avg.    0.5       0.544      0.521     0.5       0.508      0.633

=== Confusion Matrix ===

 a b   <-- classified as
 5 4 | a = yes
 3 2 | b = no
```

Slide Courtesy: http://www.cs.iastate.edu/~cs573x/bbsilab.html

Test Options

- Percentage Split (2/3 Training; 1/3 Testing)
- Cross-validation
  - Estimating the generalization error based on resampling when limited data
    - averaged error estimate.
  - Cross-fold validation (10-fold)
  - Leave-one-out (Loo)
  - Stratified

Slide Courtesy: http://www.cs.iastate.edu/~cs573x/bbsilab.html

Comparing Classification Methods

- Predictive accuracy
- Interpretability: providing insight
- Robustness: handling noisy data

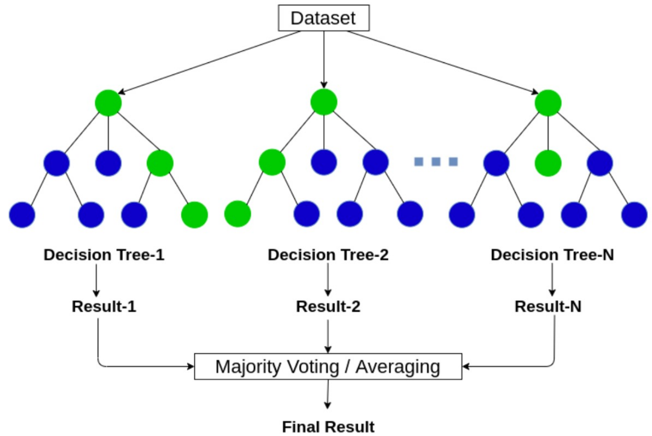
- Speed
- Scalability: large volume of data

Source: Data Mining: Concepts and Techniques, by Jiawei Han and Micheline Kamber

Classification in German Credit

- Demonstration with Weka
  - Methods to use:
    - Simple Logistic Classifier
    - Decision Tree
  - We will use 2 other methods on the same dataset soon
- Using python libraries
  - https://www.kaggle.com/sanyalush/predicting-credit-risk

Figure Credit: https://www.analyticsvidhya.com/blog/2020/05/decision-tree-vs-random-forest-algorithm/



Random Forest

- An ensemble method

- Credits
  - Ideas introduced by Tin Kam Ho in 1995, https://en.wikipedia.org/wiki/Tin_Kam_Ho
  - Matured by Leo Breiman and Adele Cutler at Berkeley (https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#intro)
  - History: Khaled Fawagreh, Mohamed Medhat Gaber & Eyad Elyan (2014) Random forests: from early developments to recent advancements, Systems Science & Control Engineering, 2:1, 602-609, DOI: 10.1080/21642583.2014.956265
  - Blog: https://www.analyticsvidhya.com/blog/2020/05/decision-tree-vs-random-forest-algorithm/

Slide Courtesy: Leo Breiman and Adele Cutler website

Random Forest

- Main steps (Input: data, N= number of trees)
  - If the number of cases in the training set is N, sample N cases at random - but *with replacement*, from the original data. This sample will be the training set for growing the tree.
  - If there are M input variables, a number m<<M is specified such that at each node, m variables are selected at random out of the M and the best split on these m is used to split the node. The value of m is held constant during the forest growing.
  - Each tree is grown to the largest extent possible. There is no pruning.
- Choice of m is implementation dependent; affects correlation between trees and their accuracy
- Characteristics:
  - Fast
  - Accurate
  - Unexplainable

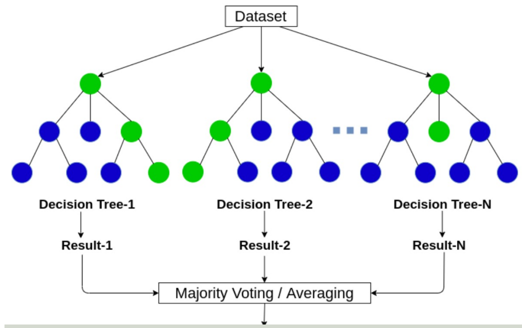
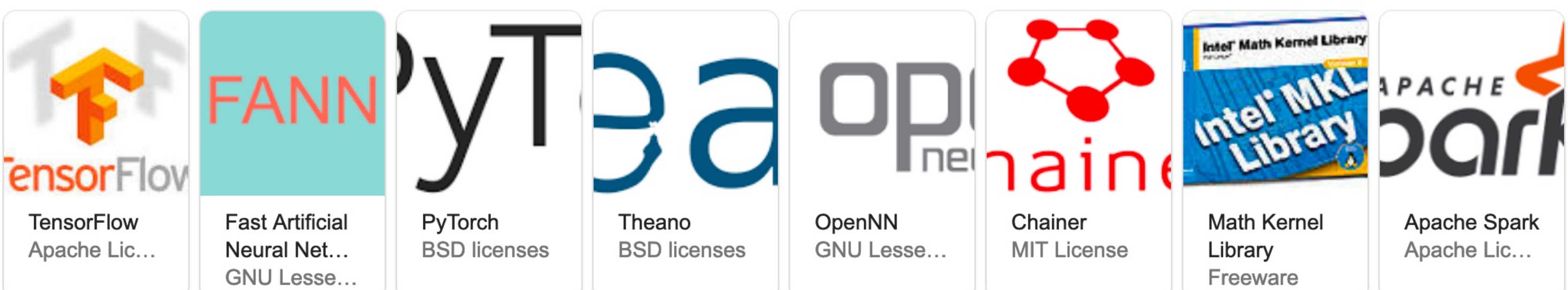


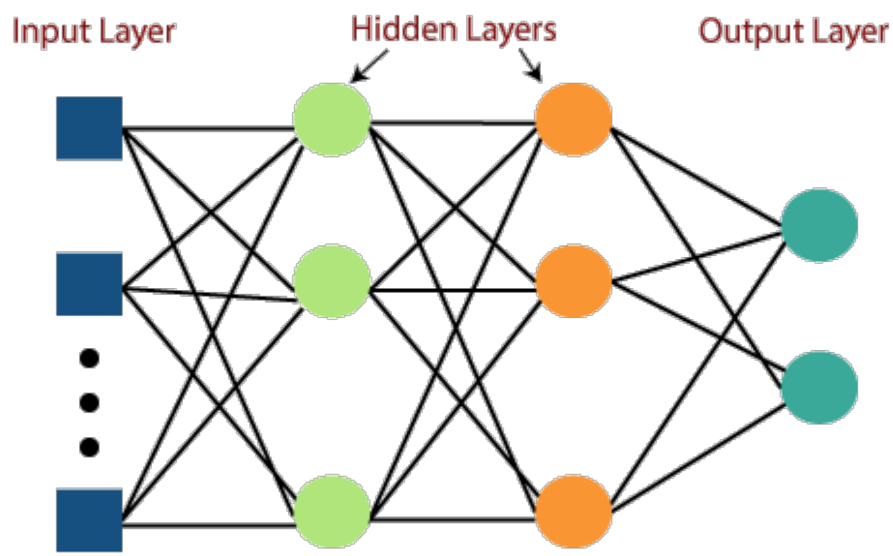
Figure Credit: https://www.analyticsvidhya.com/blog/2020/05/decision-tree-vs-random-forest-algorithm/

Slide Courtesy: Leo Breiman and Adele Cutler website

Neural Network Methods



NN – Multi Layer Perceptron



Content and Image Courtesy:
https://github.com/Thanasis1101/MLP-from-scratch

Logistic Regression in a Slide

<u>Function estimate (linear)</u>
W: weight, b: bias

$$f(X_j) = X_j W + b$$

<u>Update Weight</u>

$$W^* = W - \eta \frac{dL}{dW}$$

<u>Error Term (mean squared error)</u>

$$MSE = \frac{1}{n} \sum_{j=1}^{n} \left[f(X_{j\cdot}) - y_j \right]^2$$

<u>Common Code Pattern</u>

```
y = tf.matmul(x, W) + b
loss = tf.reduce_mean(tf.square(y - y_label))
```

NN with Keras and TensorFlow

- By Example:
  - https://github.com/biplav-s/course-nl/blob/master/l9-ml-review/Basic%20TensorFlow%20and%20Keras.ipynb
- TensorFlow's NMIST tutorial
  - https://www.tensorflow.org/tutorials/quickstart/beginner

Classification in German Credit

- Demonstration with Weka
  - Random Forest Classifier
  - Multi Layer Perceptron Classifier

- Read about more classifiers at:
  - https://machinelearningmastery.com/imbalanced-classification-of-good-and-bad-credit/
  - https://www.analyticsvidhya.com/blog/2020/05/decision-tree-vs-random-forest-algorithm/

Comparing and Choosing Supervised ML Methods

Discussion: 10 Tips Paper

- Access: https://biodatamining.biomedcentral.com/articles/10.1186/s13040-017-0155-3
- Chicco, D. Ten quick tips for machine learning in computational biology. *BioData Mining* **10,** 35 (2017). https://doi.org/10.1186/s13040-017-0155-3

The Tips

- Tip 1: Check and arrange your input dataset properly
- Tip 2: Split your input dataset into three independent subsets (training set, validation set, test set), and use the test set only once you complete training and optimization phases
- Tip 3: Frame your biological problem into the right algorithm category
- Tip 4: Which algorithm should you choose to start? The simplest one!
- Tip 5: Take care of the imbalanced data problem
- Tip 6: Optimize each hyper-parameter
- Tip 7: Minimize overfitting
- Tip 8: Evaluate your algorithm performance with the Matthews correlation coefficient (MCC) or the Precision-Recall curve
- Tip 9: Program your software with open source code and platforms
- Tip 10: Ask for feedback and help to computer science experts, or to collaborative Q&A online communities

Machine Learning – Insights from Data

- Descriptive analysis
  - Describe a past phenomenon
  - **Methods**: classification, clustering, dimensionality reduction, anomaly detection, neural methods
- Predictive analysis
  - Predict about a new situation
  - **Methods**: time-series, neural networks
- Prescriptive analysis
  - What an agent should do
  - **Methods**: simulation, reinforcement learning, reasoning

- New areas
  - Counterfactual analysis
  - Causal Inferencing
  - Scenario planning

References

- Insead course
  - Description: https://inseaddataanalytics.github.io/INSEADAnalytics/CourseSessions/Sessions67/ClassificationAnalysisReading.html
  - Data analytics for Business: https://inseaddataanalytics.github.io/INSEADAnalytics/

- Textbooks
  - Data Mining: Concepts and Techniques, by Jiawei Han and Micheline Kamber, https://hanj.cs.illinois.edu/bk3/
  - Introduction to Data Mining (Second Edition), by Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar, https://www-users.cse.umn.edu/~kumar001/dmbook/index.php

Lecture 8: Project Discussion

In Class Presentation / Review

Trust Issues

Context: German Credit Data's Analysis

- Review detailed data exploration at:
https://www.kaggle.com/sanyalush/predicting-credit-risk

- Notice issues in lending ?
  - No single female
  - Discrimination by gender, age, … ?

Discussion on Reading Material - 1

"Biases in AI Systems", Ramya Srinivasan, Ajay Chander
Communications of the ACM, August 2021, Vol. 64 No. 8, Pages 44-49
10.1145/3464903

https://cacm.acm.org/magazines/2021/8/254310-biases-in-ai-systems/fulltext

Discussion on Reading Material

"Data science and AI in the age of COVID-19 - Reflections on the response of the UK's data science and AI community to the COVID-19 pandemic", Alan Turing Institute, 2021

- Findings
  1. Researchers responded to COVID need with enthusiasm leading to a large number of projects
     1. **Word-wide* (for context)**: protein to aid disease detection and treatment (**molecular scale**), the analysis of patient data like images and conditions to improve patient care (**clinical scale**) and analysis of cases and social media to predict disease severity, understand mis-information and communicate effectively (**societal scale**).
     2. **UK specific examples**: model disease spread, navigate lockdown
  2. Major hurdle was lack of "robust and timely data", especially access and standardization
     1. Develop protocols for collecting and managing protected data
     2. Develop protocols for generating anonymized and synthetic data

* Joseph Bullock, Alexandra Luccioni, Katherine Hoffmann Pham, Cynthia Sin Nga Lam, and Miguel Luengo-Oroz. Mapping the landscape of artificial intelligence applications against covid-19. In Journal of Artificial Intelligence Research 69, 807-845, 2020.

Discussion on Reading Material

"Data science and AI in the age of COVID-19 - Reflections on the response of the UK's data science and AI community to the COVID-19 pandemic", Alan Turing Institute, 2021

- Findings
  3. Concern over inequality and exclusion slowed progress. Inadequate representation and engagement from some groups
  4. Challenge in communicating research findings to policy makers and public. Specifically, timeliness, accuracy and clarity.
     - Communication among experts
     - Communication among researchers and policy makers
     - Communication among researchers and public

* Joseph Bullock, Alexandra Luccioni, Katherine Hoffmann Pham, Cynthia Sin Nga Lam, and Miguel Luengo-Oroz. Mapping the landscape of artificial intelligence applications against covid-19. In Journal of Artificial Intelligence Research 69, 807-845, 2020.

Project Discussion

Course Project

- **Framework**
  1. (Problem) Think of a problem whose solution may benefit people (e.g., health, water, air, traffic, safety)
  2. (User) Consider how the primary user (e.g., patient, traveler) may be solving the problem today
  3. (AI Method) Think of what the solution will do to help the primary user
     1. Solution => ML task (e.g. classification), recommendation, text summarization, …
     2. Use a foundation model (e.g., LLM-based) solution as the baseline
  4. (Data) Explore the data for a solution to work
  5. (Reliability: Testing) Think of the evaluation metric we should employ to establish that the solution will works? (e.g., 20% reduction in patient deaths)
  6. (Holding Human Values) Discuss if there are fairness/bias, privacy issues?
  7. (Human-AI) Finally, elaborate how you will explain the primary user that your solution is trustable to be used by them

Project Discussion: What to Focus on ?

- Problem: you should care about it
- Data: should be available
- Method: you need to be comfortable with it. Have at least two – one serves as baseline
- Trust issue
  - Due to Users
    - Diverse demographics
    - Diverse abilities
    - Multiple human languages
  - Or other impacts
- What one does to mitigate trust issue

Rubric for Evaluation of Course Project

Project

- Project plan along framework introduced (7 points)
- Challenging nature of project
- Actual achievement
- Report
- Sharing of code

Presentation

- Motivation
- Coverage of related work
- Results and significance
- Handling of questions

Project Discussion

1. Create a private Github repository called "CSCE581-Spring2025-<studentname>-Repo". Share with Instructor (biplav-s)
2. Create a folder called "Project". Inside, create a text file called "ProjectPlan.md" (or "ProjectPlan.txt") and have details by the next class (Jan 30, 2025)

1. Title:
2. Key idea: (2-3 lines)
3. Who will care when done:
4. Data need:
5. Methods:
6. Evaluation:
7. Users:
8. Trust issue:

Concluding Section

Week 4 (L7 and L8): Concluding Comments

- We looked at
  - Supervised ML algorithms, ML tools
  - Deep-dive into German credit

- Project descriptions finalized

About Next Week – Lectures 9, 10

Lectures 9, 10: Quiz / ML Methods and Trust

- Class 9 - Quiz 1
- Classification Trust methods

5	Jan 28 (Tu)	Common AI methods: ML Landscape
6	Jan 30 (Th)	AI - Structured: Analysis – Supervised ML
7	Feb 4 (Tu)	AI - Structured: Analysis – Supervised ML
8	Feb 6 (Th)	Project discussion (1)
9	Feb 11 (Tu)	Quiz 1
10	Feb 13 (Th)	AI - Structured: Analysis – Supervised ML – Trust Issues
11	Feb 18 (Tu)	AI - Structured: Analysis – Supervised ML – Trust Issues
12	Feb 20 (Th)	AI - Structured: Analysis – Supervised ML – Mitigation Methods