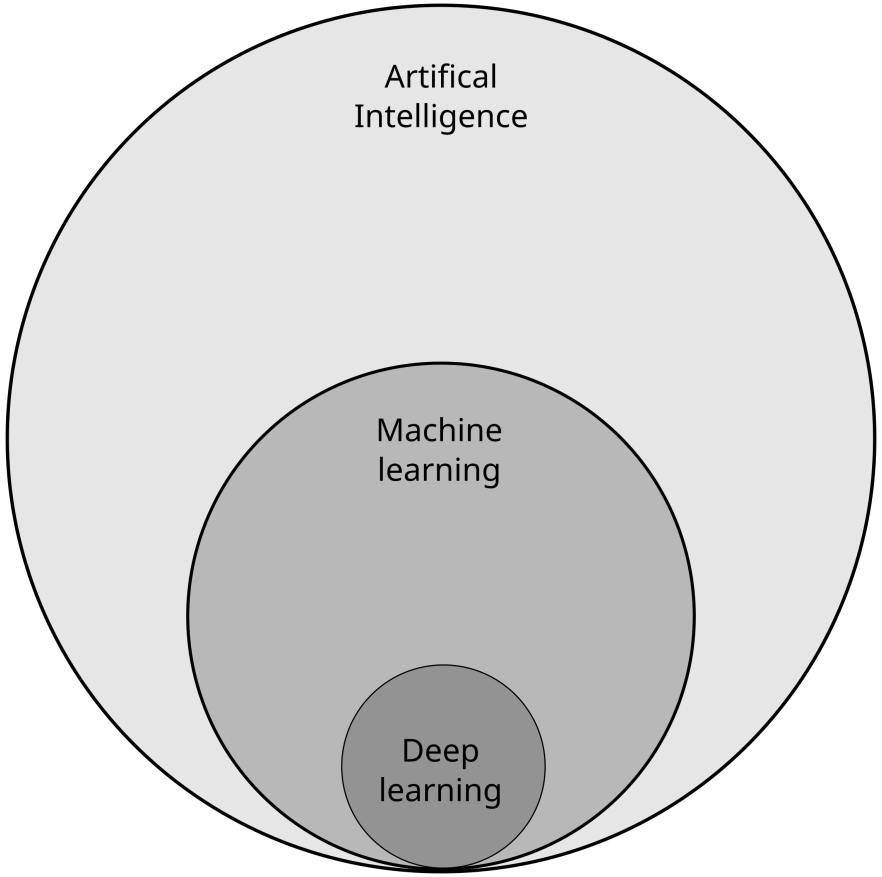
# Introduction to Health Analytics

Introduction to Machine Learning for Health Analytics

Dr. Sam Burn

# Today

# What is machine learning?



# Uses of AI in health analytics

AI is used across the healthcare system to:

# Examples of AI tools used in healthcare

| Tool | Used for | Used by |
|---|---|---|
| Aidoc | Facilitating early disease detection and minimizing diagnostic errors | Hospitals |
| Shift Technology | Streamlining claims processing and fraud detection | Insurance companies |
| Atomwise | Identifying potential drug candidates and predicting their efficacy | Drug development |
| BenevolentAI | Decreasing the time and expense associated with bringing new treatments to market | Drug development |
| Digital Diagnostics | Detecting conditions such as diabetic retinopathy and skin cancer from medical images | Diagnostics |
| PathAI | Achieving results comparable to those of experienced clinicians in diagnostics | Diagnostics |
| Qventus | Enhancing operational efficiency using AI-powered predictive analytics | Healthcare facilities |
| LeanTaaS | Optimizing resource allocation and reducing operational costs | Healthcare facilities |

*Source: DataCamp, "AI in Healthcare"*
https://www.datacamp.com/blog/ai-in-healthcare

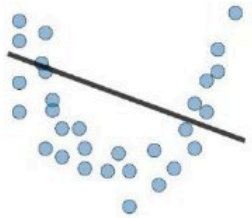
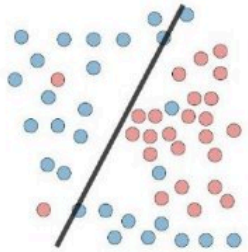
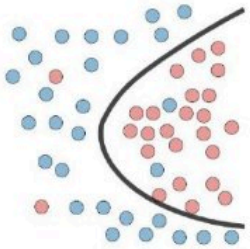
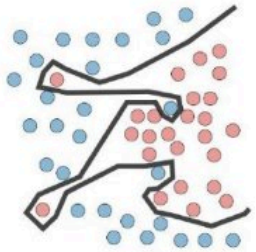
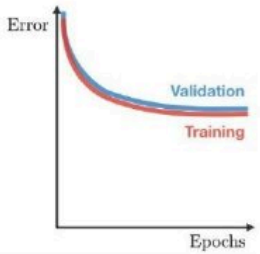
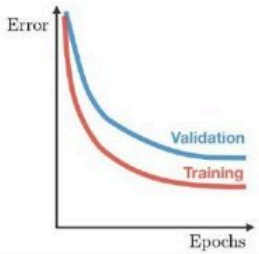
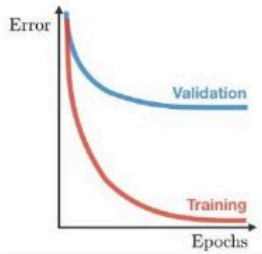
# What is the goal of machine learning?

# Training and testing data


Used for model development

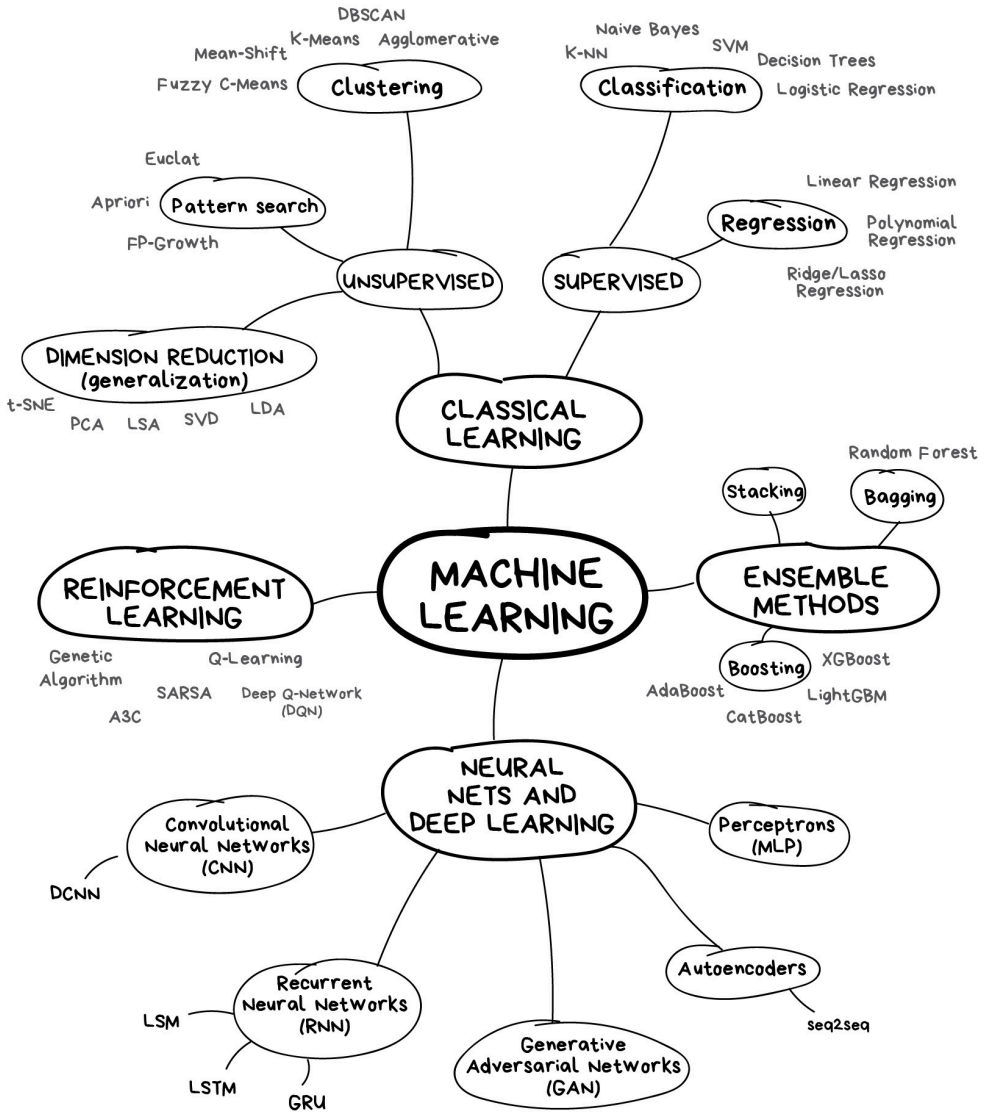
Used to evaluate final model

# The problem of overfitting

| | Underfitting | Just right | Overfitting |
|---|---|---|---|
| **Symptoms** | - High training error<br>- Training error close to test error<br>- High bias | - Training error slightly lower than test error | - Low training error<br>- Training error much lower than test error<br>- High variance |
| **Regression** |  |  |  |
| **Classification** |  |  |  |
| **Deep learning** |  |  |  |
| **Remedies** | - Complexify model<br>- Add more features<br>- Train longer | | - Regularize<br>- Get more data |

Source: Wonseok Shin

# Machine learning methods

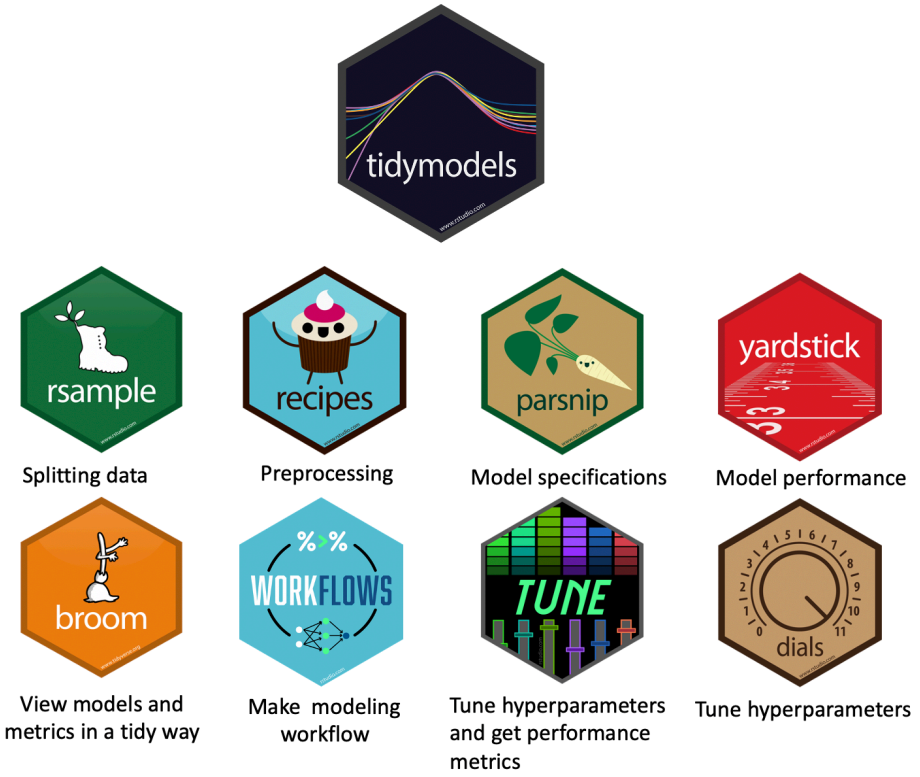



Source: https://vas3k.com/blog/machine_learning/

# A basic typology of machine learning methods

# A note on terminology

# Machine learning in R: {tidymodels}




Source: rpubs.com/chenx/tidymodels_tutorial

# Recipe and workflow

# Exercise: Exploratory data analysis and standard logit

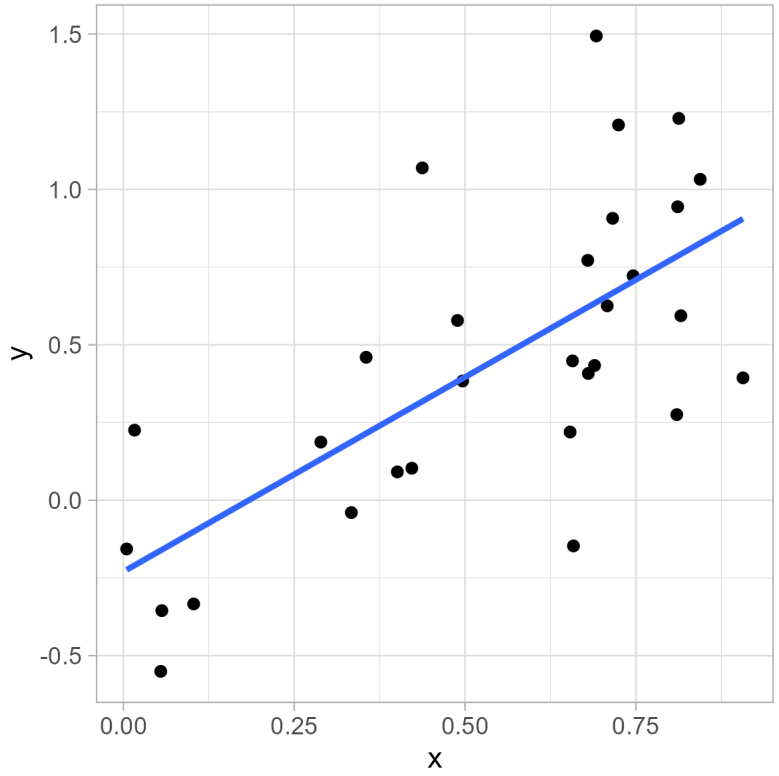
10:00

# LASSO regression

# Linear and logistic regression models

## Linear regression

```
1 lm(y ~ x, data = model_data)
```



# LASSO regression

# LASSO regression

# Hyperparameters

# Evaluating model fit

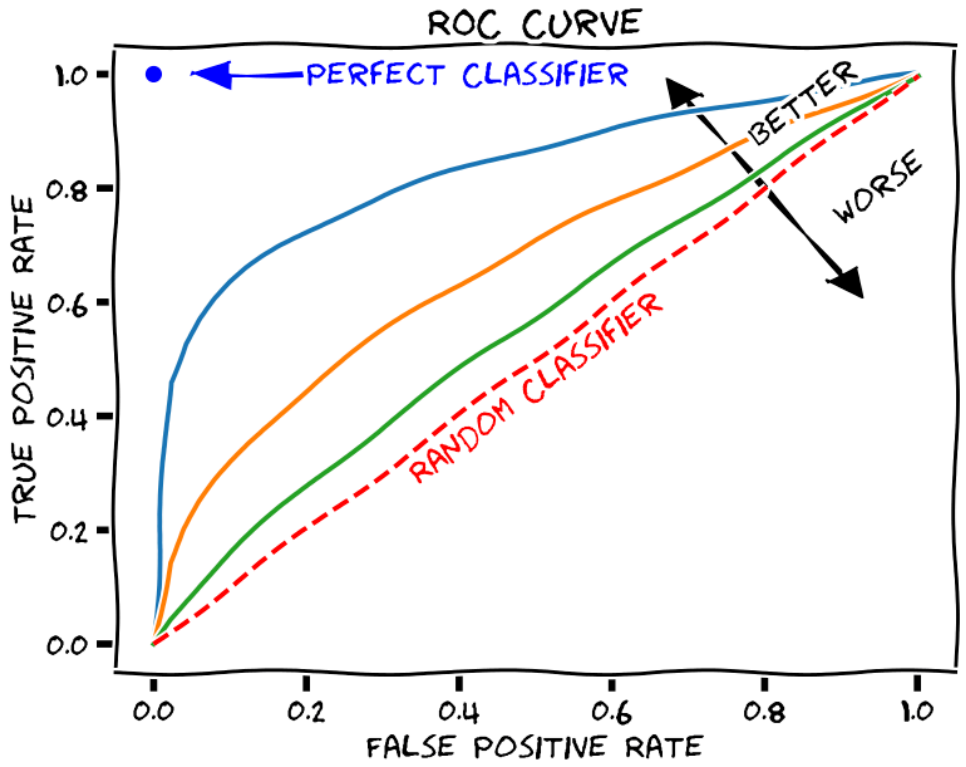
# Evaluating model fit

# Confusion matrix

| | **Predicted 1** | **Predicted 0** |
|---|---|---|
| **Actual 1** | True Positive (TP) | False Negative (FN) |
| **Actual 0** | False Positive (FP) | True Negative (TN) |

# The limitation of confusion matrices

# ROC curve: performance across thresholds




Source: Martin Thoma (Wikipedia)

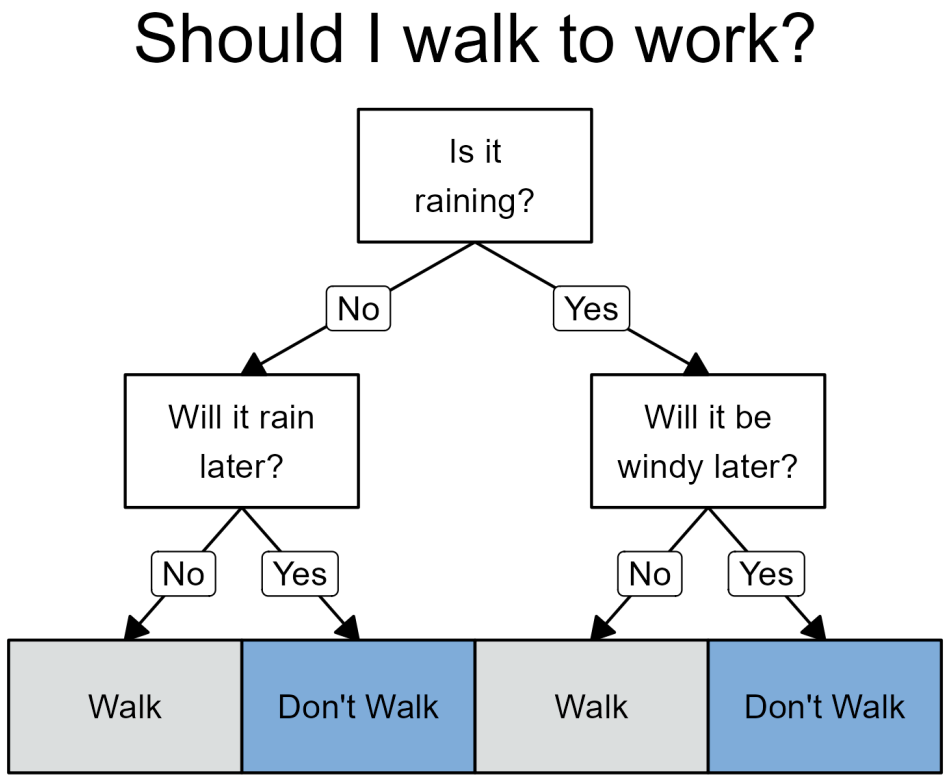
# Exercise: Estimating a LASSO logistic regression in R

10:00

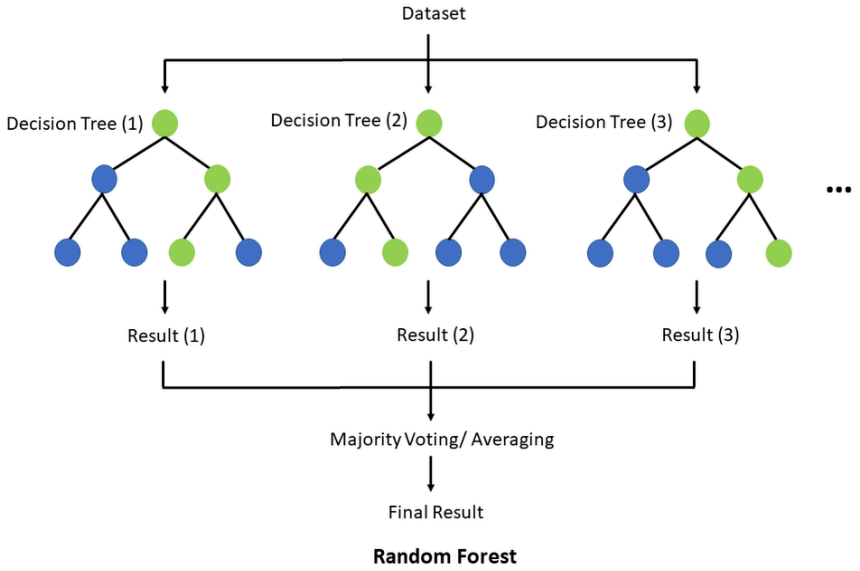
# Random Forests

# Decision trees

A tree-like model of decisions and their possible consequences.



# What are Random Forests?




Source: Tse Ki Chun (Wikimedia)

# Hyperparameters for random forests

# Exercise: Random Forests in {tidymodels}

10:00

# Takeaways

# Thank you