WORKSHOPS

Week 7 – Intro to Machine Learning

Danai Korre

MODEL SELECTION AND EVALUATION

Empirical error and overfitting

Evaluation methods

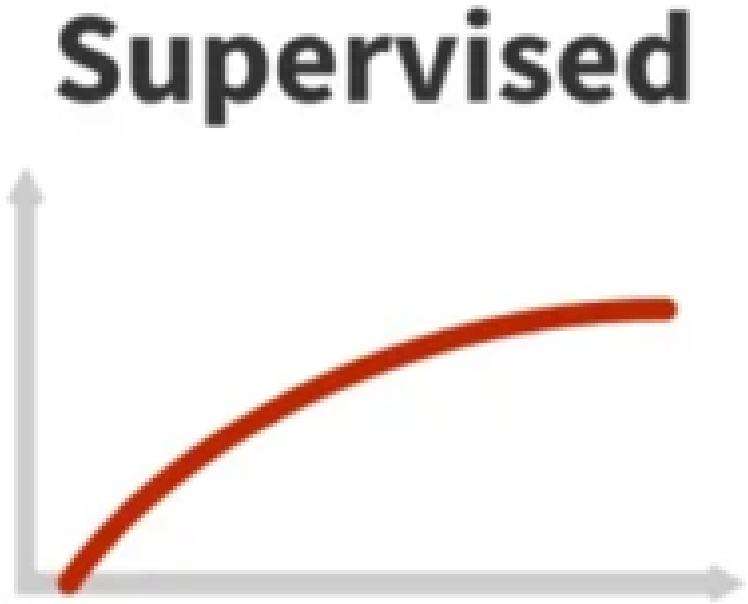
Performance measure

Comparison test

Bias and variance

SUPERVISED LEARNING

WHAT IS SUPERVISED LEARNING?



Source: Ben Freundorfer

Doug Rose defines supervised learning as "When a data scientist acts like a tutor for the machine, training it by showing it basic rules and giving it an overall strategy." [5]

- **Regression model**
- **Classification model**

SUPERVISED MACHINE LEARNING

- We humans learn from past experiences.
- A computer does not “experience.”
  - **A computer system learns from data, which** represents “past experiences” in an application domain.
- Our focus: learn a target function that can be used to predict the values (labels) of a discrete class attribute, e.g.,
  - **high-risk** or **low risk** and **approved** or **not-approved**.
- The task is commonly called: supervised learning, classification, or inductive learning.

Source: Bing Liu, UIC

EXAMPLE APPLICATION

A credit card company receives thousands of applications for new cards. Each application contains information about an applicant,

- age
- annual salary
- outstanding debts
- credit rating
- etc.

Problem: Decide whether an application should be approved, i.e., **classify** applications into two categories, **approved** and **not approved**.

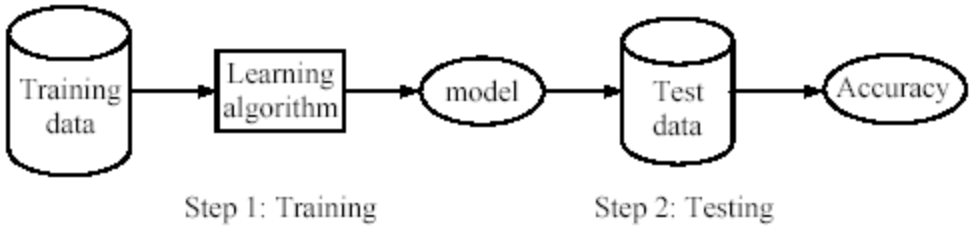
Source: Bing Liu, UIC

SUPERVISED LEARNING PROCESS: TWO STEPS

Learning or training: Learn a model using the training data (with labels)

Testing: Test the model using unseen test data (without labels) to assess the model accuracy

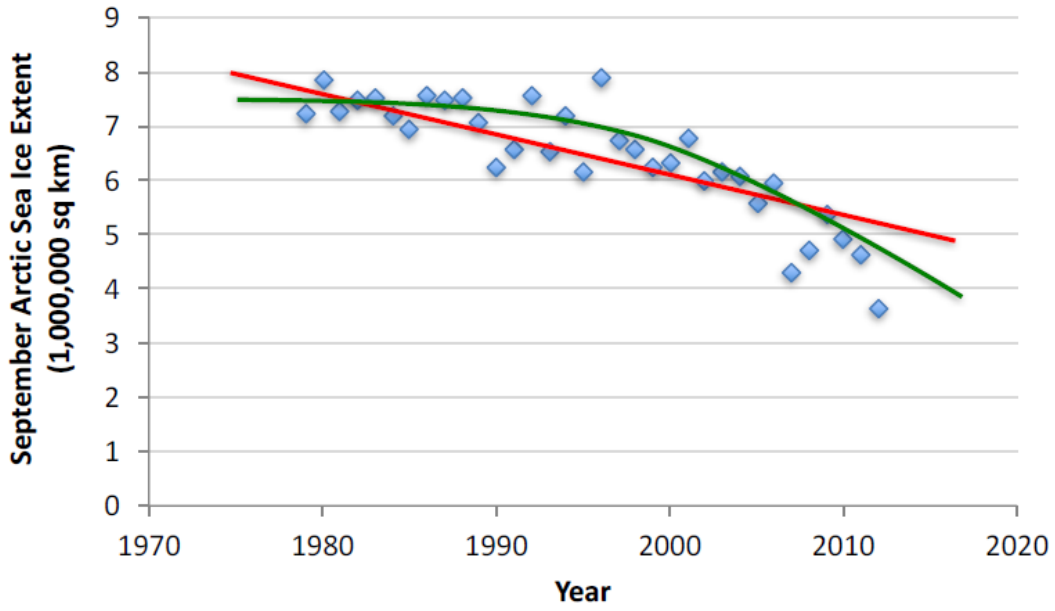
$$Accuracy = \frac{\text{Number of correct classifications}}{\text{Total number of test cases}},$$



Source: Bing Liu, UIC

REGRESSION MODEL

- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is real-valued == regression



Data from G. Witt. Journal of Statistics Education, Volume 21, Number 1 (2013)

Source: Eric Eaton

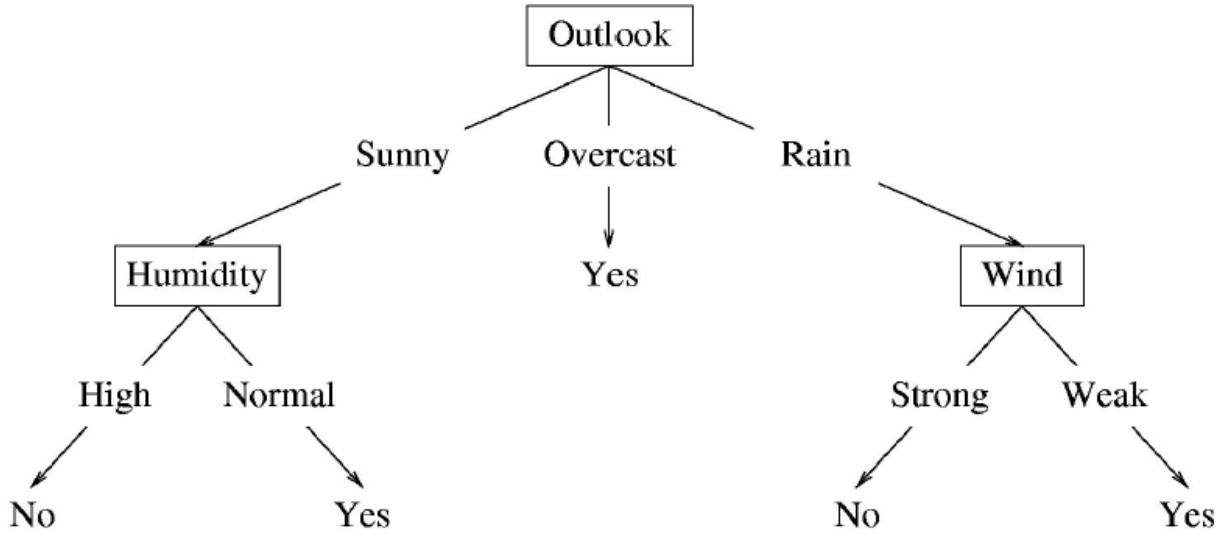
REGRESSION

Types of Regression Algorithm:

- Simple Linear Regression
- Multiple Linear Regression
- Polynomial Regression
- Support Vector Regression
- Decision Tree Regression
- Random Forest Regression

DECISION TREE

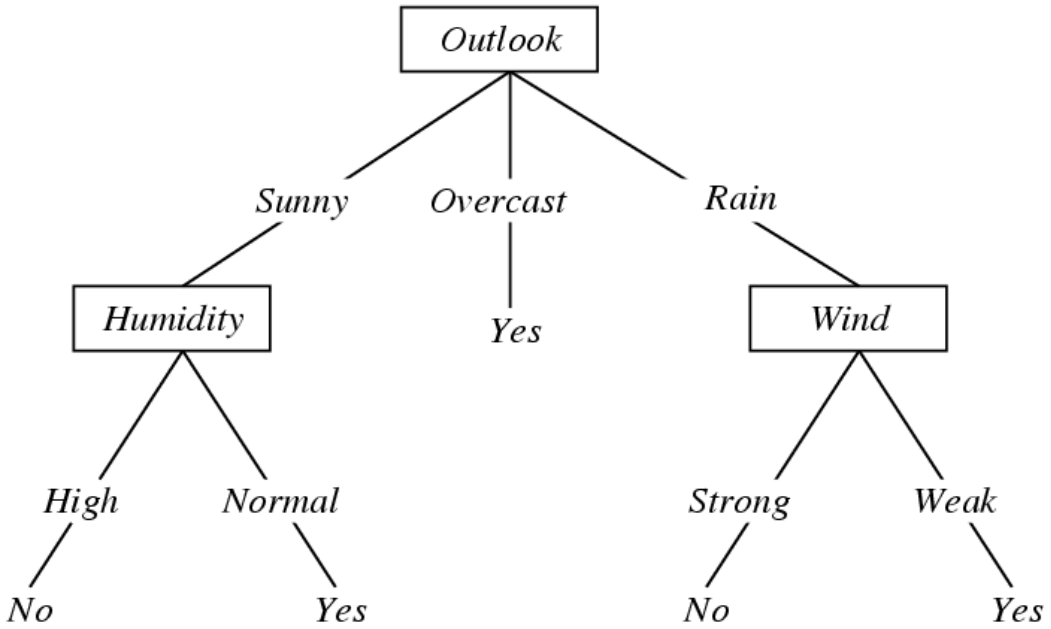
- A possible decision tree for the data:



- Each internal node: test one attribute X_i
- Each branch from a node: selects one value for X_i
- Each leaf node: predict Y (or $p(Y \mid \boldsymbol{x} \in \text{leaf})$)

DECISION TREE LEARNING

Decision Tree for *PlayTennis*

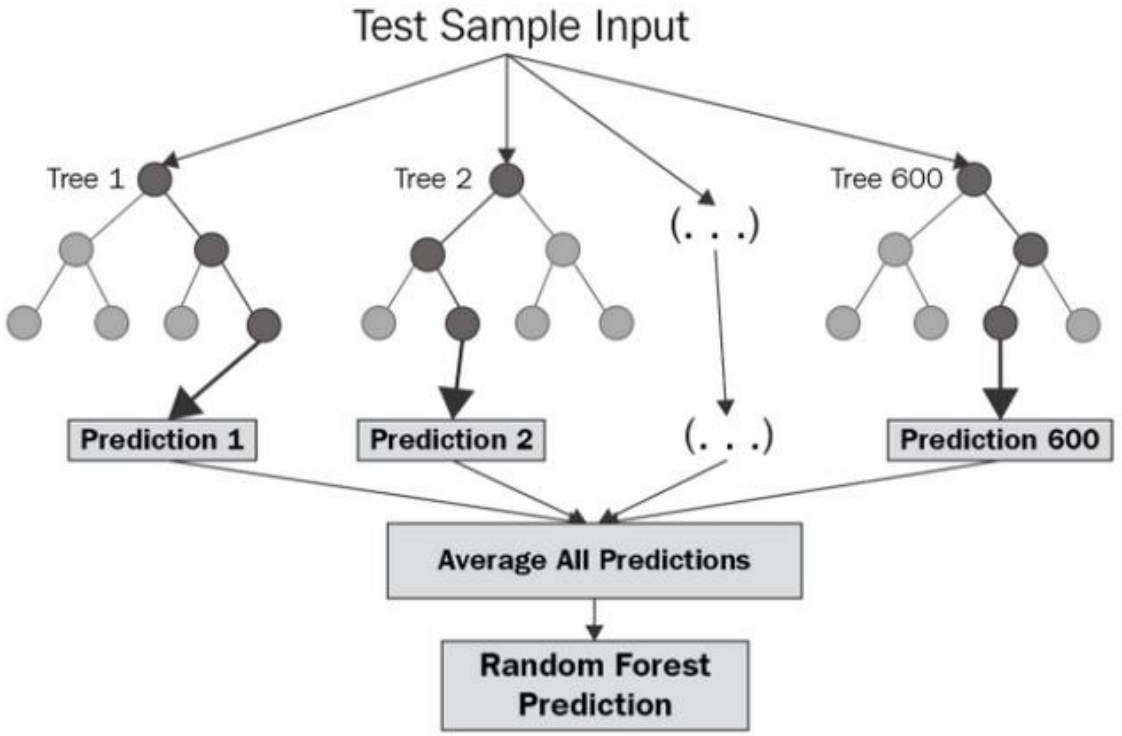


Problem Setting:

- Set of possible instances *X*
  - each instance *x* in *X* is a feature vector
  - e.g., <*Humidity=low, Wind=weak, Outlook=rain, Temp=hot*>
- Unknown target function $f : X \rightarrow Y$
  - *Y* is discrete valued
- Set of function hypotheses $H=\{ h \mid h : X \rightarrow Y \}$
- each hypothesis *h* is a decision tree
- trees sorts *x* to leaf, which assigns *y*

RANDOM FOREST REGRESSION

Random Forest Regression is a supervised learning algorithm that uses **ensemble learning** method for regression. Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model.



Source: Chaya Bakshi

RANDOM FOREST REGRESSION

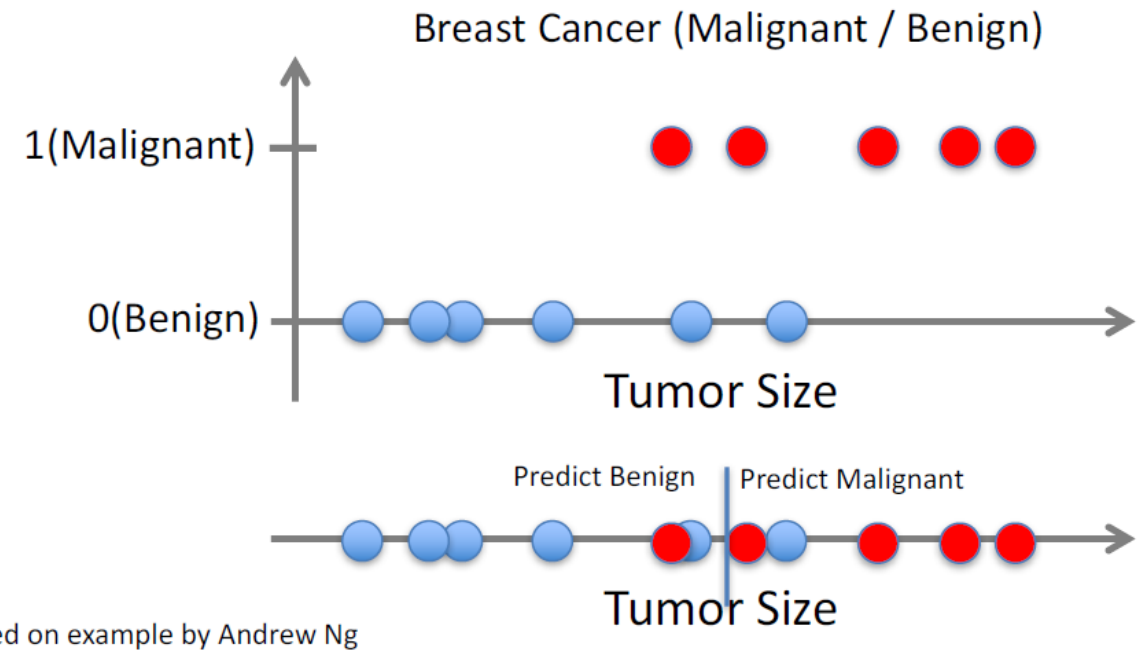
To get a better understanding of the Random Forest algorithm, let's walk through the steps:

- Pick at random k data points from the training set.
- Build a decision tree associated to these k data points.
- Choose the number N of trees you want to build and repeat steps 1 and 2.
- For a new data point, make each one of your N trees predict the value of y for the data point in question and assign the new data point to the average across all of the predicted y values.

Source: Chaya Bakshi

CLASSIFICATION MODEL

- Given (x_1, y_1), (x_2, y_2), ..., (x_n, y_n)
- Learn a function $f(x)$ to predict y given x
  - y is categorical == classification



Based on example by Andrew Ng

Source: Eric Eaton

CLASSIFICATION

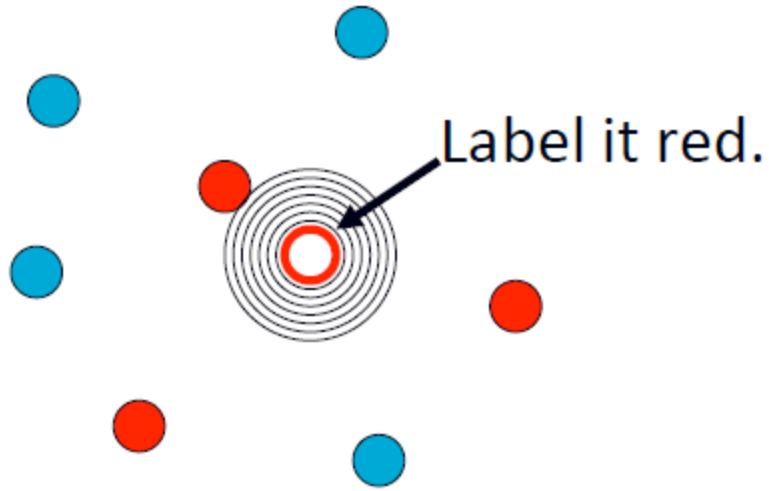
Types of ML Classification Algorithms:

Classification Algorithms can be further divided into the following types:

- Logistic Regression
- K-Nearest Neighbours
- Support Vector Machines
- Kernel SVM
- Naïve Bayes
- Decision Tree Classification
- Random Forest Classification

K-NEAREST NEIGHBOUR

- **1-Nearest Neighbour**
- One of the simplest of all machine learning classifiers
- Simple idea: label a new point the same as the closest known point



Source: Eric Eaton

K-NEAREST NEIGHBOUR

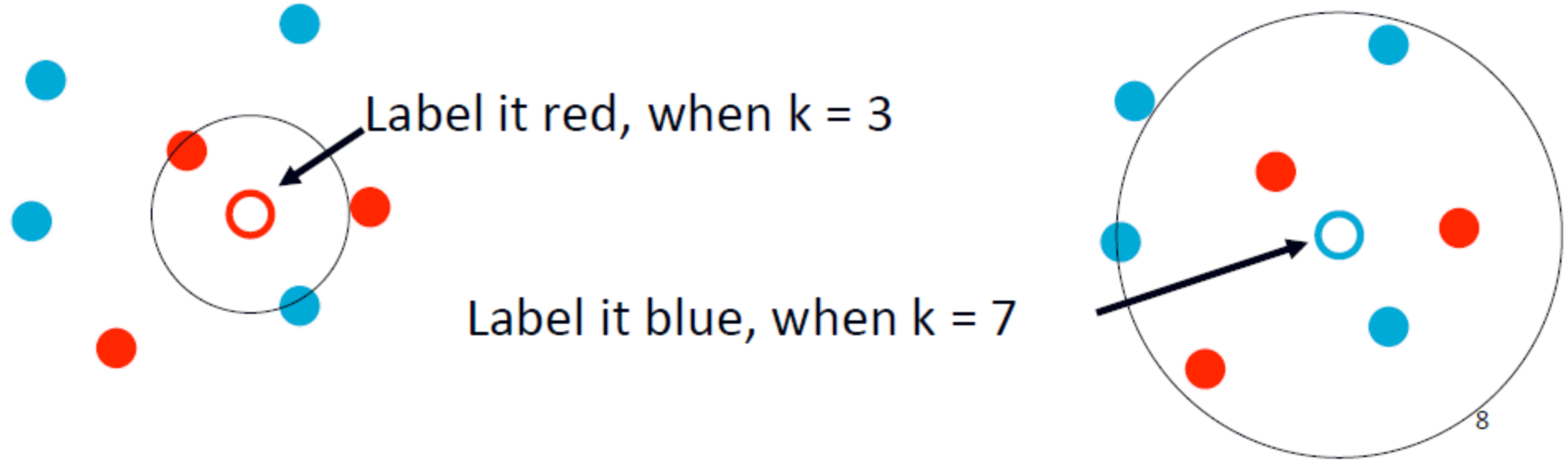
- **Four Aspects of an Instance---Based Learner:**
- 1. A distance metric
- 2. How many nearby neighbours to look at?
- 3. A weighting function (optional)
- 4. How to fit with the local points?

Adapted from "Instance---Based Learning" lecture slides by Andrew Moore, CMU.

Source: Eric Eaton

K-NEAREST NEIGHBOUR

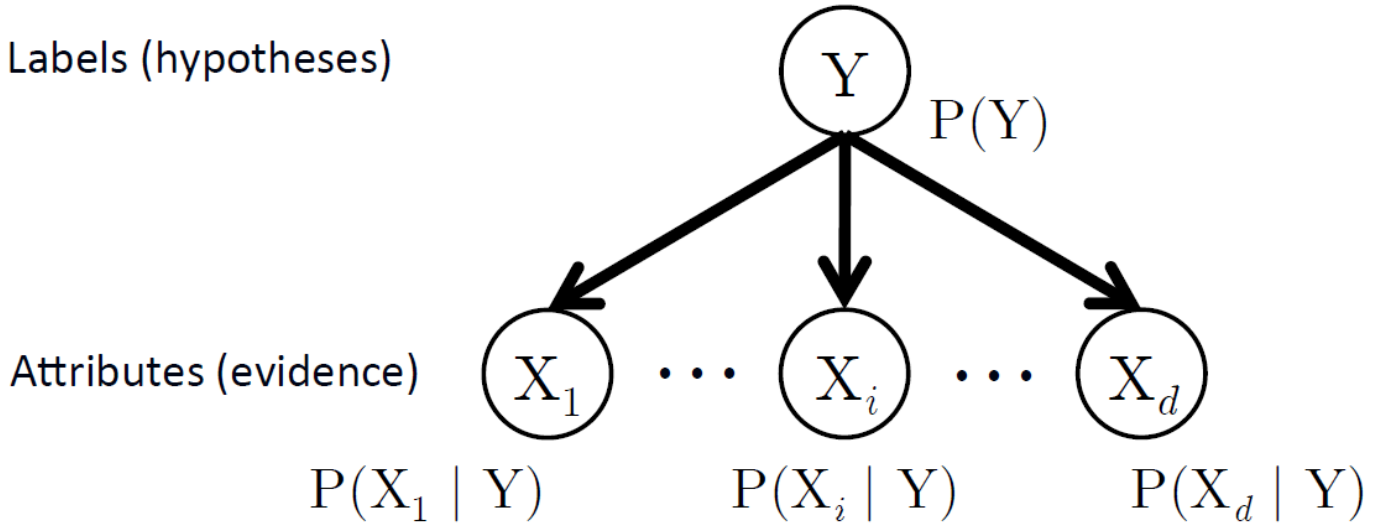
- Generalizes 1---NN to smooth away noise in the labels
- A new point is now assigned the most frequent label of its *k* nearest neighbours



Source: Eric Eaton

NAIVE BAYES

The Naïve Bayes Graphical Model

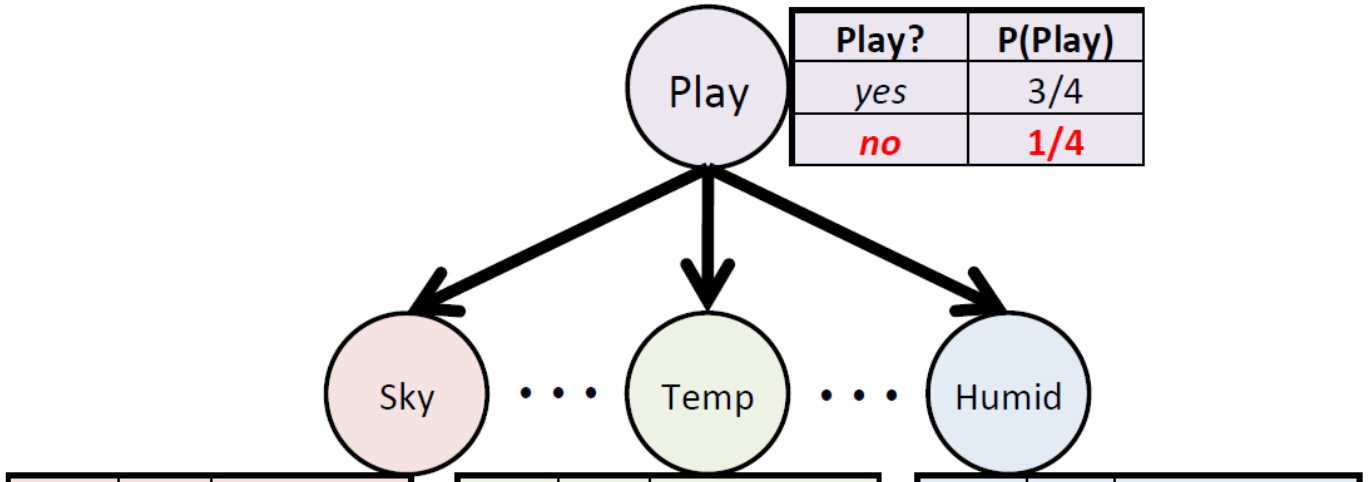


- Nodes denote random variables
- Edges denote dependency
- Each node has an associated conditional probability table (CPT), conditioned upon its parents

Source: Eric Eaton

NAIVE BAYES

Example NB Graphical Model



Play?	P(Play)
yes	3/4
no	**1/4**

Sky	Play?	P(Sky | Play)
sunny	*yes*	4/5
rainy	***yes***	**1/5**
sunny	*no*	1/3
rainy	***no***	**2/3**

Temp	Play?	P(Temp | Play)
warm	*yes*	4/5
cold	***yes***	**1/5**
warm	*no*	1/3
cold	***no***	**2/3**

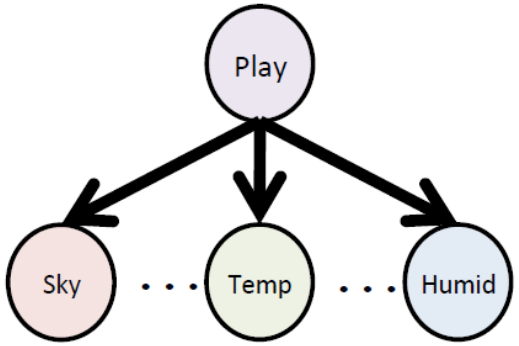
Humid	Play?	P(Humid | Play)
high	*yes*	3/5
norm	***yes***	**2/5**
high	*no*	2/3
norm	***no***	**1/3**

- Some **redundancies** in CPTs that can be eliminated

Source: Eric Eaton

NAIVE BAYES

Example Using NB for Classification



Play?	P(Play)
yes	3/4
no	1/4

Temp	Play?	P(Temp | Play)
warm	*yes*	4/5
cold	*yes*	1/5
warm	*no*	1/3
cold	*no*	2/3

Sky	Play?	P(Sky | Play)
sunny	*yes*	4/5
rainy	*yes*	1/5
sunny	*no*	1/3
rainy	*no*	2/3

Humid	Play?	P(Humid | Play)
high	*yes*	3/5
norm	*yes*	2/5
high	*no*	2/3
norm	*no*	1/3

$$h(\mathbf{x}) = \arg\max_{y_k} \log P(Y = y_k) + \sum_{j=1}^{d} \log P(X_j = x_j \mid Y = y_k)$$

Goal: Predict label for $\mathbf{x}$ = (rainy, warm, normal)

Source: Eric Eaton

NAIVE BAYES

- **Advantages:**
  - Fast to train (single scan through data)
  - Fast to classify
  - Not sensitive to irrelevant features
  - Handles real and discrete data
  - Handles streaming data well
- **Disadvantages:**
  - Assumes independence of features

REFERENCES