Artificial Intelligence Methods for Social Good

M3-1 [Machine Learning]:

Basics of Machine Learning

08-537 (9-unit) and 08-737 (12-unit)

Instructor: Fei Fang

feifang@cmu.edu

Wean Hall 4126

Outline

- What is Machine Learning
- Models
  - Decision Tree
  - Gaussian Mixture Model
  - Bayesian Network
  - Markov Random Field

  Probabilistic Graphical Models
- Evaluation
- Algorithms
  - Greedy decision tree learning
  - Maximum Likelihood Estimation (MLE)
  - Expectation-maximization (EM)

Learning Objectives

- Understand the concept of
  - Supervised learning, Unsupervised learning
  - Classification, Clustering
  - Decision Tree
  - Gaussian Mixture Model
  - Bayesian Network
  - Markov Random Field
- Understand Bayes' theorem
- List the commonly used evaluation criteria for classification task
- Describe the key ideas of
  - Greedy decision tree learning
  - Maximum Likelihood Estimation (MLE)
  - Expectation-maximization (EM)
- Know how to find the algorithm/solver/package to do classification/clustering/probabilistic inference using the models introduced

What is Machine Learning

- Definition
  - Data → Intelligence
  - Tom Mitchell: "A computer program is said to *learn* from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."
  - My favorite: Input x → (Not fully hard-coded) Program → Output y

- Applications
  - Face recognition
  - Handwriting recognition
  - Language translation
  - House price prediction
  - Automated medical diagnosis
  - Product recommendation
  - Spam filtering

What is Machine Learning

- If one proposes a ML model, he should answer the following questions
  - Representation: How to represent the relationship between input and output
  - Inference: How to infer the output from input
  - Learning: How to learn the best model to describe the data?

Input x $\rightarrow$ (Not fully hard-coded) Program $\rightarrow$ Output y

What is Machine Learning

- Know correct output for some input?
  - Supervised learning
  - Unsupervised learning
  - Semi-supervised learning

- Type of output?
  - Classification (Discrete value)
  - Regression (Real value)
  - Clustering (Group)

Supervised learning

Unsupervised learning

Outline

- What is Machine Learning
- Models
  - Decision Tree
  - Gaussian Mixture Model
  - Bayesian Network
  - Markov Random Field
- Evaluation
- Algorithms
  - Greedy decision tree learning
  - Maximum Likelihood Estimation (MLE)
  - Expectation-maximization (EM)

Decision Tree

- Widely used in the real world
- Exp1: Predict risk of loan (Safe vs Risky)
  - Credit score
  - Current income
  - Amount of loan
  - Type of loan
  - …
- A decision tree consists of
  - Decision nodes: Root node + Intermediate nodes
  - Leaf nodes / End nodes

Decision Tree

- Predict output given a hard-coded decision tree
  - Input → Traverse the tree → Output

- Learn/Train a decision tree
  - Given a set of input data and output label
  - Find a proper decision tree (Greedy Algorithm)
  - Code packages: sklearn (Python), fitctree (MATLAB)

Quiz 1: Decision Tree

- Given a spam filter represented by the decision tree shown on the whiteboard, if I send an email using email address xxx@123.com with title "Hello" and content "I love Pittsburgh", will it be classified as spam?
- Yes
- No

Decision Tree

- Beyond the example
  - Real-valued features
  - Multi-class labels
  - Regression trees

Outline

- What is Machine Learning
- Models
  - Decision Tree
  - Gaussian Mixture Model
  - Bayesian Network
  - Markov Random Field
- Evaluation
- Algorithms
  - Greedy decision tree learning
  - Maximum Likelihood Estimation (MLE)
  - Expectation-maximization (EM)

Gaussian Mixture Models

- Often used for clustering
- Exp 2: Group scene images into three groups, check average pixel intensity of Blue in each group

Gaussian Mixture Models

- Gaussian (Normal) distribution: most commonly used
  - Mean
  - Variance
  - Gaussian distribution: $\mathrm{X} \sim \mathcal{N}(\mu, \sigma^2)$
  - $f(x|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Gaussian Mixture Models

- Model: Mixture of Gaussian
  - Distribution of each class/group/cluster: Close to Gaussian
  - Aggregated distribution: Combination of weighted Gaussians
  - Weight: Relative proportion of each class

Gaussian Mixture Models

- Predict output given a hard-coded GMM
  - Input → Check probability for each group → Output
  - Check probability: Bayes' Theorem

$$P(A \mid B) = \frac{P(B|A)P(A)}{P(B)}$$

- Learn/Train GMM
  - Given a set of data
  - Find the weight, mean and variance of each Gaussian component (Expectation Maximization algorithm)
  - Code packages: sklearn (Python), fitgmdist (MATLAB)

Quiz 2: Gaussian Mixture Models

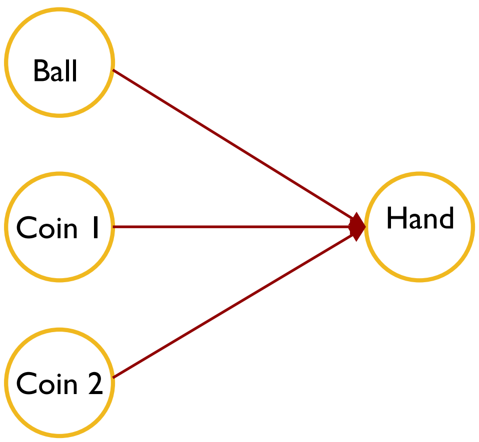
- Given a GMM for a corpus of images as follows, with the average blue intensity as the only feature used
  - 30% images belong group 1: $\mathcal{N}(0.2, 0.1)$
  - x% images belong to group 2: $\mathcal{N}(0.8, 0.05)$
  - y% images belong to group 3: $\mathcal{N}(0.5, 0.1)$
  - x and y are unknown
  - For an image whose average blue intensity is 0.7, which group does it belong to?
    - Group 1
    - Group 2
    - Group 3
    - Cannot be determined

Outline

- What is Machine Learning
- Models
  - Decision Tree
  - Gaussian Mixture Model
  - Bayesian Network
  - Markov Random Field
- Evaluation
- Algorithms
  - Greedy decision tree learning
  - Maximum Likelihood Estimation (MLE)
  - Expectation-maximization (EM)

Bayesian Network

- Recall Q1, HW0: A bag had three balls in it, two red and one blue. Alice randomly picked a ball from the bag and then checked its color. Now Bob asked Alice: "What is the color of your ball?" Alice told him: "I will flip two normal coins. I will raise my left hand if I get two heads and my ball is red or if I get at least one tail and my ball is blue. Otherwise, will raise my right hand." Bob saw Alice raised her right hand. What is the probability that Alice's ball is blue (up to two digits after decimal point)?



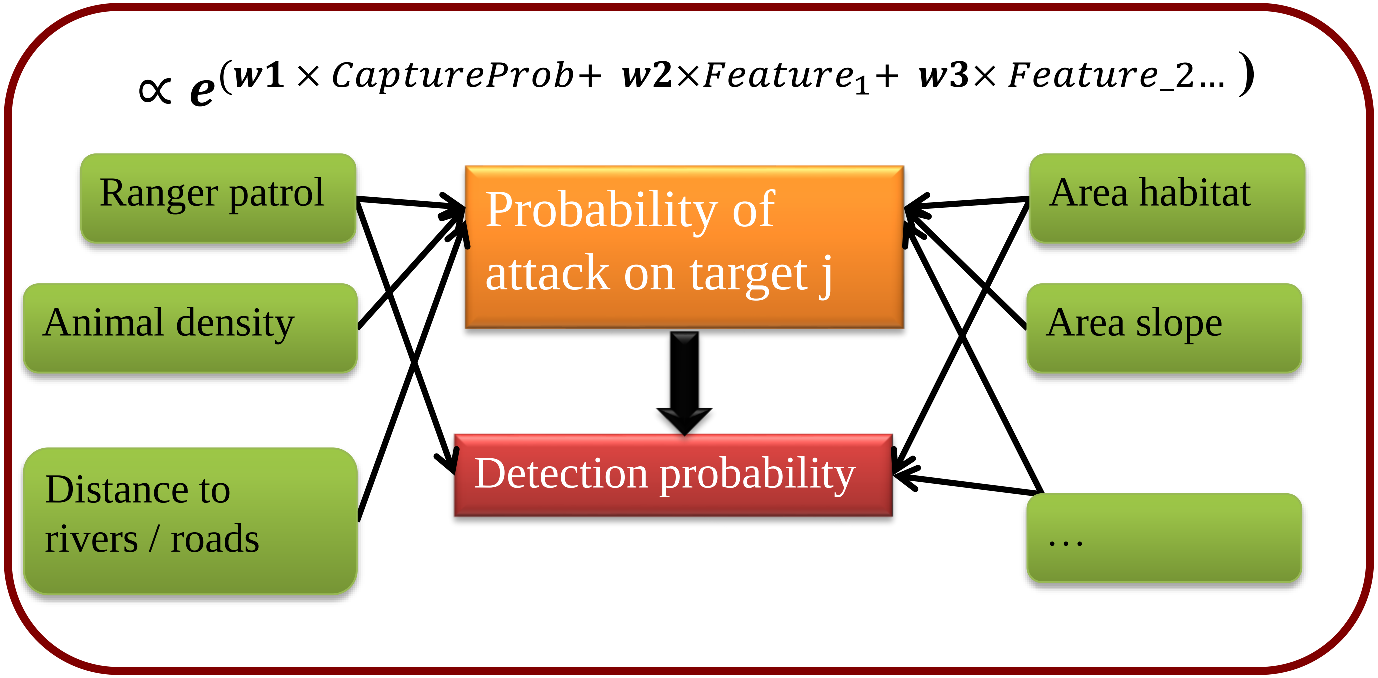
Bayesian Network

- Bayesian Network
  - Directed acyclic graph
  - Nodes: random variables
  - Edges: dependency relationship
  - Each node is associated with a conditional probability distribution: combination of parent values → probability
  - Joint probability = apply chain rule on network

- Exercise: Build the Bayesian Network for Genetic Inheritance

Bayesian Network

- Exp 3: Recall: CAPTURE for estimating poaching probabilities



Bayesian Network

- Predict output given a hard-coded BN
  - Input → Check probability for each option → Output
  - Check probability: Bayes' Theorem

- Learn/Train BN
  - Given a set of data
  - Find network structure and the conditional probability distribution of each node (Constraint-based, Scoring-based)
  - Code packages: sklearn, bayespy (Python), boot.strength + bn.fit (R)

Bayesian Network

- Interpret Gaussian Mixture Model as a special class of Bayesian Network

- Hidden Markov Model
  - An important special class of Bayesian Network
  - An extension to Gaussian Mixture Model
  - State transition
  - State $\rightarrow$ observation
  - Learn/Train HMM: Dynamic Programming (Viterbi algorithm)

Bayesian Network

- Mimic how the data is generated
- Dependency relationship (Not necessarily causality)

Outline

- What is Machine Learning
- Models
  - Decision Tree
  - Gaussian Mixture Model
  - Bayesian Network
  - Markov Random Field
- Evaluation
- Algorithms
  - Greedy decision tree learning
  - Maximum Likelihood Estimation (MLE)
  - Expectation-maximization (EM)

Markov Random Field

- When there is no causality-like dependency, only correlation

- Markov Random Field
  - Undirected graph
  - Nodes: random variables
  - Edges: correlation relationship
  - Each clique is associated with a potential function: combination of node values in the clique → score
  - Joint probability = normalized multiplication of potential function

- Exercise: Build the Markov Random Field for an image of handwritten digit

Markov Random Field

- Predict output given a hard-coded MRF
  - Input → Check probability for each option → Output
  - Check probability: compute joint probability based on potential functions, apply Bayes' Theorem (intractable in general)

- Learn/Train MRF
  - Given a set of data
  - Find the parameters of potential functions for each clique (e.g., Maximum Likelihood Estimation)
  - Code packages: PyStruct (Python)

Outline

- What is Machine Learning
- Models
  - Decision Tree
  - Gaussian Mixture Model
  - Bayesian Network
  - Markov Random Field
- Evaluation
- Algorithms
  - Greedy decision tree learning
  - Maximum Likelihood Estimation (MLE)
  - Expectation-maximization (EM)

Evaluation (Classification)

- Split the data into training set and test set
- Learn/Train a model using training set
- Test the model using test set
- Report performance based on evaluation metrics

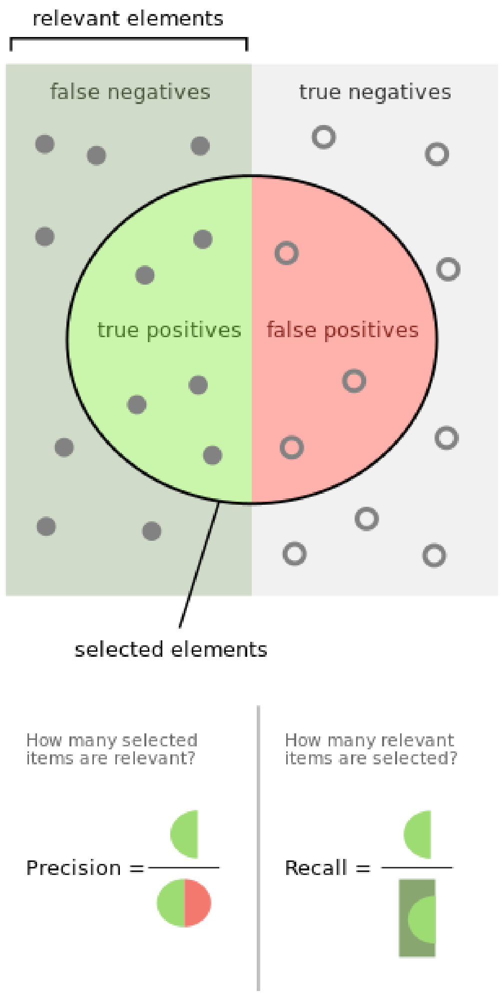
Evaluation (Classification)

- Basic metrics
  - Error
  - Accuracy

- Alice has a rarely used email address and she has found that 90% of the emails sent to this address are spam. She asks you to design a spam filter for her. You designed one that labels every email as spam. What is the expected accuracy of this classifier?

Evaluation (Classification)

- Commonly used metrics
  - Confusion matrix
  - Precision
  - Recall
  - F1 score: harmonic mean of precision and recall



Quiz 3: Evaluation (Classification)

- Alice has a rarely used email address and she has found that 90% of the emails sent to this address are spam. She asks you to design a spam filter for her. You designed one that labels every email as spam. What is the F1 score of this classifier (keep two digits after the decimal point)?
- 0.90
- 0.10
- 0.95
- 0.85

Outline

- What is Machine Learning
- Models
  - Decision Tree
  - Gaussian Mixture Model
  - Bayesian Network
  - Markov Random Field
- Evaluation
- Algorithms
  - Greedy decision tree learning
  - Maximum Likelihood Estimation (MLE)
  - Expectation-maximization (EM)

Algorithm: Greedy Decision Tree Learning

- Step 1: Start with an empty tree
- Step 2: Split on a feature
- Step 3: Making predictions
- Go back to Step 2

Algorithm: Greedy Decision Tree Learning

- Problem 1: Feature split selection
  - Greedily choose the feature to minimize error when assigning majority label
- Problem 2: Stopping condition
  - 1) All data in the node have same y value
  - 2) Already split on all features

Maximum Likelihood Estimation

- Find parameters to maximize the *likelihood* of data
- Example: Given samples from a normal distribution, estimate mean and variance

- How?
  - Special case (e.g., Gaussian): Known formula
  - General case: gradient descent
  - Solvers: fminunc (MATLAB), scipy (Python)

Expectation Maximization

- Goal: estimate parameters in a model (e.g., mean and variance of GMM)
- Divide variables into two groups X, Y
- E-step
  - Estimate value/distribution of X given current estimates of parameters and value/distribution of Y
- M-step
  - Maximize likelihood over parameters given current values of X and Y

Expectation Maximization

- Example: Gaussian Mixture Models
- E-step: estimate probability of belonging to each cluster given current parameter estimates
- M-step: maximize likelihood over parameters given current cluster probabilities

Summary

- Basic Concepts in Machine Learning

- Key take-away:
  - Learn to perform task through experience
  - Evaluation metric is important
  - An ML model: Representation, Inference, Learning

Additional Resources

- Text book
  - *Pattern Recognition and Machine Learning*, *Chapters 4, 8, 9*
  - Christopher Bishop
- Online course
  - https://www.coursera.org/specializations/probabilistic-graphical-models
  - https://www.coursera.org/learn/ml-clustering-and-retrieval
  - https://www.coursera.org/learn/ml-classification
  - https://www.coursera.org/learn/machine-learning