quiz 4

Probability

Axioms of probability:

- 0 <= P(a) <= 1
- P(NOT(a)) = 1 – P(a)
- P(true) = 1
- P(false) = 0
- P(A OR B) = P(A) + P(B) – P(A AND B)

Concepts of probability

And Probability

$$P(A, B) = P(A^\wedge B) = P(A) + P(B) - P(A^\vee B)$$

Or Probability

$$P(A^\vee B) = P(A) + P(B) – P(A^\wedge B)$$

Conditional Probability

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

Product Rule

- aka **Chain Rule**
- $P(a, b) = P(a|b)P(b) = P(b|a)P(a)$

Using Product Rule

$$P(a, b, c) = P(a, b|c)P(c) = P(a|b, c)P(b, c)$$
$$P(a, b, c|d, e) = P(a|b, c, d, e)P(b, c|d, e)$$

Sum Rule

- aka **Law of Total Probability**
  $P(A) = \sum_{B,C} P(A, B, C)$
  e.g.
  $P(b) = \sum_a \sum_c \sum_d P(a, b, c, d)$
  $P(a, d) = \sum_b \sum_c P(a, b, c, d)$
- Given a set of probabilities $P(CatchFish, Day, Lake)$
- Where:
  - CatchFish = {true, false}
  - Day = {mon, tues, wed, thurs, fri, sat, sun}
  - Lake = {buel lake, ralph lake, crystal lake}
    – Need to find P(CatchFish = True):
  - $P(CatchFish = true) = \Sigma_{day} \Sigma_{lake} P(CatchFish = true, day, lake)$

Bayes' Rule

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

Derivation of Bayes' Rule

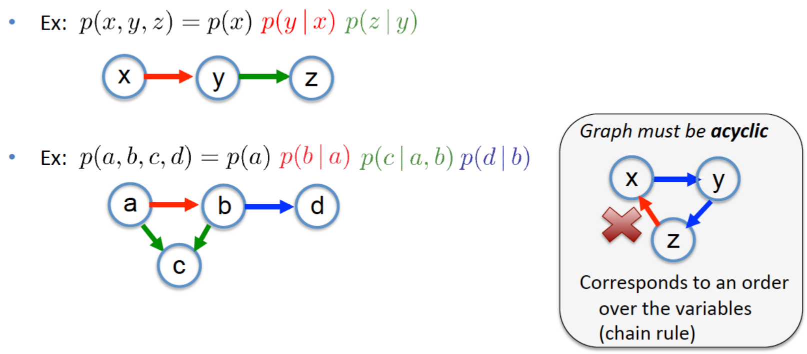
- start from Product Rule:
  - $P(a, b) = P(a|b)P(b) = \frac{P(b|a)P(a)}{P(b)} \cdot P(b) = P(b|a)P(a) = \frac{P(b,a)}{P(a)} \cdot P(a) = P(a, b)$
  - $P(a, b) = P(b|a)P(a) = \frac{P(a|b)P(b)}{P(a)} \cdot P(a) = P(a|b)P(b) = \frac{P(a,b)}{P(b)} \cdot P(b) = P(a, b)$
- Isolate Equality on Right Side:
  - $P(a|b)P(b) = P(b|a)P(a) = \frac{P(a,b)}{P(a)} \cdot P(a) = P(a, b) = \frac{P(a,b)}{P(b)} \cdot P(b) = P(a|b) \cdot P(b)$
- Divide through by P(b)
  - $P(a|b) = P(b|a)P(a)/P(b) = \frac{P(a,b)}{P(a)} \cdot P(a)/P(b) = \frac{P(a,b)}{P(b)}$

Conditional Independence

- X, Y independent given Z
- $p(X = x, Y = y|Z = z) = p(X = x|Z = z)p(Y = y|Z = z)$
  for all x,y,z
- **Equivalent**: $p(X|Y, Z) = p(X|Z) or p(Y|X, Z) = p(Y|Z)$

- **Intuition**: X has no additional info about Y beyond Z's

Bayesian Networks



Machine Learning

Important Concepts

- **Learning**: Improves performance of future tasks after observing the world
- **Information Gain**: Expected reduction in entropy from testing an attribute value
- **Decision Boundary**: Surface in a high-dimensional space that separates the classes
- **Cross-validation**: Randomly split the data into a training set and a test set
- **Linear Classifier**: Tests $w \cdot f > 0$, where w is a weight vector and f is a feature vector
- **Factored Representation (Feature Vector)**: Fixed set, list, or vector of features/ attributes paired with a value
- **Supervised Learning**: Agent observes input-output pairs & learns to map input to output
- **Test Set**: Examples distinct from training set, used to estimate accuracy
- **Naïve Bayes Classifier**: Tests $P(C) \prod_i P(X_i | C)$ Where C is a class label and X_i are features
- **Classification**: Supervised learning with a discrete set of possible output values
- **Decision Tree**: Internal nodes test a value of an attribute, leaf nodes=class labels
- **Regression**: Supervised learning with numeric output values
- **Training Set**: Example input-output pairs, from which to discover a hypothesis
- **Unsupervised Learning**: Agent learns patterns in the input with no explicit feedback
- **Overfitting**: Choose an over-complex model based on irrelevant data patterns

Decision Tree

A	B	A xor B
F	F	F
F	T	T
T	F	T
T	T	F