10601 Machine Learning

September 21, 2011

Mladen Kolar

Announcements

- Fill in a survey at
  http://www.surveymonkey.com/s/J7Q5JHL

- Recitations
  - Wed 6 – 7pm (this time) or
  - Fri 4 – 5 pm

- Homework 1 is due on Monday, noon

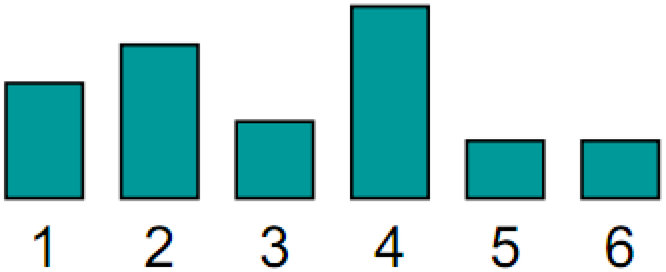
Outline

- Probability overview
- Maximum likelihood estimation

Probability review

X is discrete random variable (e.g. 6-sided die)

Probability mass function (pmf)



$P[X = x_i]$

Cumulative distribution function (cdf)

$$P[X \leq a] = \sum_{x \leq a} P[X = x]$$

Example

Role two fair 5-sided dice with labels {-2,-1,0,1,2}

- What is the pmf for the product of the two dice?
- What is the cdf for the product of the two dice?
- Let Y be 1 for a positive product, -1 for negative product and 0 for a zero product. What is the pmf for Y?

Law of total probability

Given two discrete random variables X and Y

X takes values in $\{x_1, \ldots, x_m\}$

Y takes values in $\{y_1, \ldots, y_n\}$

$P(X = x_i)$

$$P(Y = y_j) = \sum_j P(X = x_i, Y = y_j)$$??

$$P(Y = y_j) = \sum_j P(X = x_i \mid Y = y_j) P(Y = y_j)$$??

$P(X = x_i)$

Law of total probability

Given two discrete random variables X and Y

X takes values in $\{x_1, \ldots, x_m\}$

Y takes values in $\{y_1, \ldots, y_n\}$

$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$$

$$P(X = x_i) = \sum_j P(X = x_i \mid Y = y_j) P(Y = y_j)$$

Law of total probability

Given two discrete random variables X and Y

Joint probability of X,Y

$$P(X = x_i) = \sum_j \overline{P(X = x_i, Y = y_j)}$$

$$= \sum_j \underline{P(X = x_i \mid Y = y_j)} P(Y = y_j)$$

Marginal probability

Conditional probability of X conditioned on Y

Marginal probability

Defining a probability distribution

Two discrete random variables X and Y take binary values

- How do we specify a probability distribution?

Joint probabilities

$$P(X=0, Y=1) = 0.2$$

Should sum up to 1

$$P(X=0, Y=0) = 0.2$$

$$P(X=1, Y=0) = 0.5$$

$$P(X=1, Y=1) = 0.5$$

Defining a probability distribution

Two discrete random variables X and Y take binary values

– How do we specify a probability distribution?

Joint probabilities

$$P(X = 0, Y = 1) = 0.2$$

$$P(X = 0, Y = 0) = 0.2$$

$$P(X = 1, Y = 0) = 0.3$$

$$P(X = 1, Y = 1) = 0.3$$

Defining a probability distribution

Two discrete random variables X and Y take binary values

– How do we specify a marginal distribution?

Joint probabilities

$$P(X=0, Y=1) = 0.2$$

$$P(X=0, Y=0) = 0.2$$

$$P(X=1, Y=0) = 0.3$$

$$P(X=1, Y=1) = 0.3$$

Marginal probabilities

~~$P(X=0) = 0.2$~~
~~$P(X=1) = 0.8$~~

??

$$P(Y=0) = 0.5$$
$$P(Y=1) = 0.5$$

Defining a probability distribution

Two discrete random variables X and Y take binary values

- How do we specify a marginal distribution?

Joint probabilities

$$P(X=0, Y=1) = 0.2$$

$$P(X=0, Y=0) = 0.2$$

$$P(X=1, Y=0) = 0.3$$

$$P(X=1, Y=1) = 0.3$$

Marginal probabilities

$$P(X=0) = 0.4$$
$$P(X=1) = 0.6$$

$$P(Y=0) = 0.5$$
$$P(Y=1) = 0.5$$

Conditional probabilities

What is the complementary event of $P(X=0|Y=1)$?

$P(X=1|Y=1)$ OR $P(X=0|Y=0)$

Expectation

X is a discrete random variable

$$E[X] = \sum_x xP[X = x]$$

Expectation of a function

$$E[f(X)] = \sum_x f(x)P[X = x]$$

Linearity of expectation

$$E[aX + bY] = aE[X] + bE[Y]$$

Expectation (examples)

Y = flip of a fair coin (1 = heads, 0 = tails)

E[Y] = 1*P[Y = heads] + 0 * P[Y = tails] = 0.5

Z = outcome of a roll of a die

$$E[Z] = \sum_{z=1}^{6} zP[Z = z] = \sum_{z=1}^{6} \frac{z}{6} = 3.5$$

Variance

X is a discrete random variable

$$\text{Var}[X] = \text{E}[X\text{–}\text{E}[X]]^2 = E[X^2] - E[X]^2$$

X = flip of a fair coin (1 = heads, 0 = tails)

$\text{Var}[X] = E[X^2] - \text{E}[X]^2 = \sum_{x \in \{0,1\}} x^2 P[X = x] - 0.5^2 = 0.25$

Independence

Two random variables X and Y are independent if and only if

$$P[X = x, Y = y] = P[X = x]P[Y = y], \forall x, y.$$

Knowledge of one variable does not help us to make inference about the other variable.

Independence (cont.)

X and Y are independent.

Show that

- $P[X \mid Y] = P[X]$
- $P[\neg X \mid Y] = P[\neg X]$
- $P[Y \mid X] = P[Y]$

Independence (cont.)

Liked movie	Slept	raining	P
1	1	1	0.05
1	0	1	0.1
0	0	1	0.025
0	1	1	0.075
1	1	0	0.15
1	0	0	0.3
0	0	0	0.075
0	1	0	0.225

P[slept] = 0.5

P[slept|rain = 1] = 0.125 / 0.25 = 0.5

Slept and rain are independent

Conditional independence

Two random variables may become independent when conditioned on the third variable.

$$P[X, Y|Z] = P[X|Z]P[Y|Z]$$

X and Y are conditionally independent given Z

Number of parameters

Assume X and Y take Boolean values {0,1}

How many independent parameters do you need to fully specify:

- marginal probability of X?

  P(X=0) 1 parameter only [because P(X=1)+P(X=0)=1]

- the joint probability of $P(X,Y)$?

  P(X=0,Y=0), P(X=0,Y=1), P(X=1, Y=1) 3 parameters

- the conditional probability of $P(X|Y)$?

  P(X=0|Y=0), P(X=0|Y=1) 2 parameters

- the joint probability of P(X,Y), X and Y are independent

  P(X=0), P(Y=0) 2 parameters

KL divergence

The Kullback-Leibler (KL) divergence is a measure of distance between two probability distributions.

Let p(x) and q(x) be two pmfs.

$$KL(p,q) = \sum_{x} p(x) \log \frac{p(x)}{q(x)}$$

KL divergence (cont.)

Relationship to the mutual information

$$I[X;Y] = KL(P[X,Y], P[X]P[Y])$$

Can be used to measure independence of variables.

Maximum likelihood estimation

$$\hat{\theta} = \arg\max \quad P(D|\theta)$$

$$\hat{\theta} = \arg\max \quad \ln P(D|\theta)$$

Poisson distribution

$$P[X = x|\theta] = \frac{\theta^x e^{-\theta}}{x!}, \qquad x = 0,1,2, \ldots$$

Maximum likelihood estimation (cont.)

$X_1, X_2, \ldots, X_n$ are independent random variables with the Poisson distribution.

$$P[D|\theta] = \frac{\theta^{x_1} e^{-\theta}}{x_1!} \times \cdots \times \frac{\theta^{x_n} e^{-\theta}}{x_n!}$$

$$P[D|\theta] = \frac{\theta^{\sum_i x_i} e^{-n\theta}}{x_1!\, x_2! \ldots x_n!}$$

Joint distribution function (as a function of D)

Likelihood function (as a function of θ)

$$\ln P[D|\theta] = \left(\sum x_i\right) \log \theta - n\theta - \sum(\log x_i!)$$

Maximum likelihood estimation (cont.)

Log-likelihood function

$$l(\theta; D) = \ln P[D|\theta] = (\sum x_i) \log \theta - n\theta - \sum(\log x_i!)$$

$$\frac{dl}{d\theta} = \theta^{-1} \sum x_i - n = 0$$

$$\hat{\theta} = \frac{\sum x_i}{n} = \bar{x}$$ Maximum likelihood estimate

$$\frac{d^2 l}{d\theta^2} = -\frac{\sum x_i}{\theta^2} < 0$$ Stationary point is the maximum

Questions?