Announcements

HW2

- Due Sat 10/2

Quizzes

- Mon 10/4, last 15 min. of class (calculus, optimization, Lagrange)
- Mon 10/11, last 15 min. of class (probability, statistics)

Mathematical Foundations for Machine Learning

Probability

Instructor: Pat Virtue

Plan

Today

Probability

- Probability exercises
- Chain rule
- Independence
- i.i.d.

Statistics

- Finding the best parameters

Probability

Poll 1: Exercise

Implement a function in Python for the pdf of a Gaussian distribution.

Python numpy or math packages are fine, no scipy, etc.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Probability Vocab

Outcomes

Sample space

Events

Probability

Random variable

Discrete random variable

Continuous random variable

Probability mass function

Probability density function

Parameters

Probability Toolbox

- Algebra
- Three axioms of probability
- Theorem of total probability
- Definition of conditional probability
- Product rule
- Bayes’ theorem
- Chain rule
- Independence
- Conditional independence

Tools Summary

Adding to our toolbox

1. Definition of conditional probability $$P(A|B) = \frac{P(A,B)}{P(B)}$$
2. Product Rule $$P(A,B) = P(A|B)P(B)$$
3. Bayes' theorem $$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$
4. Chain Rule…

The Chain Rule

More generally, can always write any joint distribution as an incremental product of conditional distributions

Independence

Conditional Independence

Probability Notation

Notation doc

Independent and identically distributed

Likelihood

Likelihood: The probability (or density) of random variable Y taking on value y given the distribution parameters, $\boldsymbol{\theta}$.

Likelihood

Likelihood: The probability (or density) of random variable Y taking on value y given the distribution parameters, $\boldsymbol{\theta}$.

Grades

Gaussian PDF: $p\left(y \mid \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$

Likelihood

Trick coin: comes up heads only 1/3 of the time

1 flip: H probability: $\frac{1}{3}$

2 flips: H,H probability: $\frac{1}{3} \cdot \frac{1}{3}$

3 flips: H,H,T probability: $\frac{1}{3} \cdot \frac{1}{3} \cdot \left(1 - \frac{1}{3}\right)$

But why can we just multiply these?

Likelihood and i.i.d

Likelihood: The probability (or density) of random variable Y taking on value y given the distribution parameters, $\boldsymbol{\theta}$.

i.i.d.: Independent and identically distributed

Bernoulli Likelihood

Bernoulli distribution:

$$Y \sim Bern(\phi) \qquad p(y \mid \phi) = \begin{cases} \phi, & y = 1 \\ 1 - \phi, & y = 0 \end{cases}$$

What is the likelihood for three i.i.d. samples, given parameter ϕ:

$$\mathcal{D} = \{y^{(1)} = 1, y^{(2)} = 1, y^{(3)} = 0\}$$

$$\prod_{i=1}^{N} p(Y = y^{(i)} \mid \phi)$$

$$= \phi \cdot \phi \cdot (1 - \phi)$$

Poll 2

Assume that exam scores are drawn independently from the same Gaussian (Normal) distribution.

Given three exam scores 75, 80, 90, which pair of parameters is a better fit (a higher likelihood)?

A) Mean 80, standard deviation 3

B) Mean 85, standard deviation 7

Use a calculator/computer.

Gaussian PDF: $p(y \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$