COS 424: Interacting with Data

Lecturer: David Blei Lecture # 3
Scribe: Anthony Soroka, Ilya Tsinis 2/12/08

I. Monty Hall Problem

1/3 chance - picked correctly intitially (don't switch), 2/3 chance - picked incorrectly initially (switch)

C_i = indicator that the car is behind door i H_{ij} = indicator that the host chooses door j when the player chooses door i

$P(H_{ij}|C_k = 1) = 0 if i = j, = 0 if j = k, = 1/2 if i = k, = 1 if i \neq k, j \neq k$(technically, also$i \neq j$)

Monty opens door 3

P($C_1|H_{13}) \alpha p(C_1) * P(H_{13}|C_1 = 1) = 1/3 * 1/2 = 1/6$

P($C_2|H_{13}) \alpha p(C_2) * P(H_{13}|C_2 = 1) = 1/3 * 1 = 1/3$

Alternate Method

X= indicator that the correct door is picked initially

$P(X = 1|$host opens a door$) = P(X = 1,$host opens a door$)\overline{P(}$host opens a door)

$P(X = 1,$host opens a door$) = P($host opens a door$|X = 1) * P(X = 1) = 1/3$

$P($host opens a door$) = 1$

Therefore, $P(X = 1|$host opens a door$) = \frac{1/3}{1} = 1/3$ So the contestant should switch

II. Probability

Continuous R.V.s

Density $p(x)$ $\int_{-\infty}^{\infty} p(x)dx = 1$

Probability is an integral over a smaller interval

$P(X\epsilon(-2.4, 6.5)) = \int_{-2.4}^{6.5} p(x)dx$

Gaussian Distribution

$P(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi * \sigma}} * e^{-(x-\mu)^2/2\sigma^2}$

Are μ ,σ^2 parameters or random variables? This is a great debate between Bayesian and Frequentists -In this class, we'll be both!

$\mu\epsilon R, \sigma^2\epsilon R^+$

Expectaion

Consider a function of an r.v. f(X) Expectation is weighted average of f(X)

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

continuous case:

E[f(X)] $= \int p(x)f(x)dx$

$\mu = E[X]$

$\sigma^2 = E[X^2] - (E[X])^2 Conditional Expectation$

E[f(X) — Y = y] $= \sum_x p(x|y)f(x)$

Units: $E[f(X)|Y = y]$ - scaler, $E[f(X)|Y]$ - random variable

Iterated Expectation (Tower Property)

$$\text{E[E[f(X) — Y = y]]} = \textstyle\sum_y p(y)E[f(X)|Y = y] \tag{1}$$
$$= \sum_y p(y) \sum_x p(x|y)f(x) \tag{2}$$
$$= \sum_y \sum_x p(x, y)f(x) \tag{3}$$
$$= \sum_y \sum_x p(x)p(y|x)f(x) \tag{4}$$
$$= \sum_x p(x)f(x) \tag{5}$$
$$= E[f(x)] \tag{6}$$

Probability Models

- Use probability as a model of observed data - Pretend that data is drawn from an unknown distribution - INFER properties of that distribution - Use our inferences for something

IID Assumption - Independent and indetically distributed - Parameter index a distribution

e.g. coin flip has Bernouli

$p(x|\pi) = \pi^{1(X=H)}(1-\pi)^{1(X=T)}$

Suppose we flip the coin N times and record the outcomes

$X_1, ..., X_n$

Likelihood Function

$p(X_1, ..., X_n \text{given} \pi) = \prod_{n=1)^N \pi^{1(X_n=H)}(1-\pi)^{1(X_n=T)}}$

log-likelihood

$L(\pi, X_i, ..., X_n) = \sum_{n=1}^N 1(X_n = H)log\pi + 1(X_n = T)log(1-\pi)$

L$(\pi, X_i, ..., X_n) = n_H log\pi + n_T log(1-\pi)$

(MLE) Maximum Likelihood Estimate (i.e. Why do we care about log-likelihood?)

The value of the parameter that maximizes the log likelihood (equivalently the likelihood) of the observed data

MLE $\hat{\pi} = \frac{1}{N}\sum_{n=1}^N 1[X_n = H] = \frac{n_h}{N}$

Why do we like MLE?

- Consistent - If we see more and more coin flips we will get closer and closer to the true probabilities