CS 112: Computer System Modeling Fundamentals

Prof. Jenn Wortman Vaughan

May 10, 2011

Lecture 12

Quiz #3

Reminders & Announcements

- Homework 4 will be a programming assignment – we will cover the algorithm that you need to implement in class this Thursday

Types of Inference

Hypothesis testing: Decide which of two or more hypotheses is more likely to true based on some data.

- Determine whether an email containing a particular set of words is more likely to be spam or not spam
- Given a student’s test score, decide if he studied or not

Types of Inference

Hypothesis testing: Decide which of two or more hypotheses is more likely to true based on some data.

- Determine whether an email containing a particular set of words is more likely to be spam or not spam
- Given a student's test score, decide if he studied or not

Parameter estimation: Model is fully specified except some unknown parameters we need to estimate.

- Estimate the bias of a coin from a sequence of flips
- Estimate the fraction of the population who prefers candidate A to candidate B based on polling data

Hypothesis Testing

Let D be the event that we observed some particular data

- D = event that I observed an email containing the words “ca$h” and “viagra”

Hypothesis Testing

Let D be the event that we observed some particular data

- D = event that I observed an email containing the words "ca$h" and "viagra"

Let $H_1, \ldots, H_k$ be disjoint and exhaustive events representing hypotheses we are choosing among

- H_1 = event that the email is spam
- H_2 = event that the email is not spam

Hypothesis Testing

Let D be the event that we observed some particular data

- D = event that I observed an email containing the words "ca$h" and "viagra"

Let H_1, …, H_k be disjoint and exhaustive events representing hypotheses we are choosing among

- H_1 = event that the email is spam
- H_2 = event that the email is not spam

What is the most likely hypothesis given the data?

Multiple Options

- The maximum likelihood (ML) hypothesis is the hypothesis that makes the data most likely

$$H^{ML} = \mathrm{argmax}_i \; P(D \mid H_i)$$

Multiple Options

- The maximum likelihood (ML) hypothesis is the hypothesis that makes the data most likely

$$H^{ML} = \text{argmax}_i \; P(D \mid H_i)$$

- The maximum a posteriori (MAP) hypothesis is the hypothesis with the maximum posterior probability

$$H^{MAP} = \text{argmax}_i \; P(H_i \mid D) = \text{argmax}_i \; P(D \mid H_i) \, P(H_i)$$

Multiple Options

- The maximum likelihood (ML) hypothesis is the hypothesis that makes the data most likely

$$H^{ML} = \text{argmax}_i \boxed{P(D \mid H_i)}$$

- The maximum a posteriori (MAP) hypothesis is the hypothesis with the maximum posterior probability

$$H^{MAP} = \text{argmax}_i\, P(H_i \mid D) = \text{argmax}_i \boxed{P(D \mid H_i)}\, P(H_i)$$

Both require an estimate of $P(D \mid H_i)$

Multiple Options

- The maximum likelihood (ML) hypothesis is the hypothesis that makes the data most likely

$$H^{ML} = \text{argmax}_i \; P(D \mid H_i)$$

- The maximum a posteriori (MAP) hypothesis is the hypothesis with the maximum posterior probability

$$H^{MAP} = \text{argmax}_i \; P(H_i \mid D) = \text{argmax}_i \; P(D \mid H_i) \boxed{P(H_i)}$$

MAP additionally requires $P(H_i)$
(can be a *subjective probability*)

Maximum Likelihood

There are two boxes of cookies. One contains half chocolate chip cookies and half oatmeal raison cookies. The other contains one third chocolate chip cookies and two thirds oatmeal raison. I select a box and pull a random cookie from it. You observe that the cookie is chocolate chip.

Which box is most likely to be the one I chose from?

Maximum Likelihood

There are two boxes of cookies. One contains half chocolate chip cookies and half oatmeal raison cookies. The other contains one third chocolate chip cookies and two thirds oatmeal raison. I select a box and pull a random cookie from it. You observe that the cookie is chocolate chip.

Which box is most likely to be the one I chose from?

- D = event that I chose a chocolate chip cookie
- $P(D \mid H_1) = 0.5$, $P(D \mid H_2) = 0.33$
- $H^{ML} = H_1$

MAP

There are two boxes of cookies. One contains half chocolate chip cookies and half oatmeal raison cookies. The other contains one third chocolate chip cookies and two thirds oatmeal raison. I select a box and pull a random cookie from it. You observe that the cookie is chocolate chip.

If you know that box 2 is on the table and box 1 is put away, which box is most likely to be the one I chose from?

MAP

There are two boxes of cookies. One contains half chocolate chip cookies and half oatmeal raison cookies. The other contains one third chocolate chip cookies and two thirds oatmeal raison. I select a box and pull a random cookie from it. You observe that the cookie is chocolate chip.

If you know that box 2 is on the table and box 1 is put away, which box is most likely to be the one I chose from?

- $P(H_1) = 0.1$, $P(H_2) = 0.9$ (for example..)
- $H^{MAP} = H_2$

Next Couple Weeks…

- Classical statistical inference
  - Parameter estimation with maximum likelihood
  - Bias, confidence bounds, other desirable properties

- Bayesian statistical inference
  - Using priors and posteriors
  - MAP estimation

- Example application: Naive Bayes classifier (which you will implement for homework 4!)

Parameter Estimation

Suppose that we would like to estimate the unknown bias p of a coin based on observations of the outcomes $X_1, \ldots, X_n$ of n independent tosses of the coin

Parameter Estimation

Suppose that we would like to estimate the unknown bias p of a coin based on observations of the outcomes $X_1, \ldots, X_n$ of n independent tosses of the coin

(This is just like our polling question…)

Parameter Estimation

Suppose that we would like to estimate the unknown bias p of a coin based on observations of the outcomes $X_1, \ldots, X_n$ of n independent tosses of the coin

(This is just like our polling question…)

We can define analogs of both ML and MAP here

Parameter Estimation

- The maximum likelihood (ML) estimate is the parameter value that makes the data most likely

$$\hat{\theta} = \arg\max_{\theta} P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n; \theta)$$

Parameter Estimation

- The maximum likelihood (ML) estimate is the parameter value that makes the data most likely

$$\hat{\theta} = \arg\max_{\theta} P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \boxed{; \theta)}$$

"parameterized by θ"

Parameter Estimation

- The maximum likelihood (ML) estimate is the parameter value that makes the data most likely

$$\hat{\theta} = \arg\max_{\theta} P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n; \theta)$$

Parameter Estimation

- The maximum likelihood (ML) estimate is the parameter value that makes the data most likely

$$\hat{\theta} = \arg\max_{\theta} P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n; \theta)$$

- If $X_1, \ldots, X_n$ are independent observations, then

$$\hat{\theta} = \arg\max_{\theta} \prod_{i=1}^{n} P(X_i = x_i; \theta)$$

Parameter Estimation

- The maximum likelihood (ML) estimate is the parameter value that makes the data most likely

$$\hat{\theta} = \arg\max_{\theta} \boxed{P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n; \theta)}$$

- If $X_1, \ldots, X_n$ are independent observations, then

$$\hat{\theta} = \arg\max_{\theta} \boxed{\prod_{i=1}^{n} P(X_i = x_i; \theta)}$$

"likelihood"

Parameter Estimation

- The maximum likelihood (ML) estimate is the parameter value that makes the data most likely

$$\hat{\theta} = \arg\max_{\theta} P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n; \theta)$$

- If $X_1, \ldots, X_n$ are independent observations, then

$$\hat{\theta} = \arg\max_{\theta} \prod_{i=1}^{n} P(X_i = x_i; \theta)$$

$$= \arg\max_{\theta} \sum_{i=1}^{n} \log\big(P(X_i = x_i; \theta)\big)$$

Parameter Estimation

- The maximum likelihood (ML) estimate is the parameter value that makes the data most likely

$$\hat{\theta} = \arg\max_{\theta} P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n; \theta)$$

- If $X_1, \ldots, X_n$ are independent observations, then

$$\hat{\theta} = \arg\max_{\theta} \prod_{i=1}^{n} P(X_i = x_i; \theta)$$

$$= \arg\max_{\theta} \boxed{\sum_{i=1}^{n} \log\left(P(X_i = x_i; \theta)\right)}$$

"log likelihood"

Parameter Estimation

Suppose that we would like to estimate the unknown bias p of a coin based on observations of the outcomes $X_1, \ldots, X_n$ of n independent tosses of the coin

What is the maximum likelihood estimate?

Parameter Estimation

Suppose that the time it takes for a certain model of hard drive to fail is an exponential random variable

We would like to estimate the unknown parameter λ based on n independent observations $X_1, \ldots, X_n$

What is the maximum likelihood estimate?

Parameter Estimation

- The maximum likelihood (ML) estimate is the parameter value that makes the data most likely

$$\hat{\theta} = \arg\max_{\theta} P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n; \theta)$$

Parameter Estimation

- The maximum likelihood (ML) estimate is the parameter value that makes the data most likely

$$\hat{\theta} = \arg\max_{\theta} f_{X_1,\ldots,X_n}(x_1, x_2, \ldots, x_n; \theta)$$

Parameter Estimation

- The maximum likelihood (ML) estimate is the parameter value that makes the data most likely

$$\hat{\theta} = \arg\max_{\theta} f_{X_1,\ldots,X_n}(x_1, x_2, \ldots, x_n; \theta)$$

- If $X_1, \ldots, X_n$ are independent observations, then

$$\hat{\theta} = \arg\max_{\theta} \prod_{i=1}^{n} P(X_i = x_i; \theta)$$

Parameter Estimation

- The maximum likelihood (ML) estimate is the parameter value that makes the data most likely

$$\hat{\theta} = \arg\max_{\theta} f_{X_1,\dots,X_n}(x_1, x_2, \dots, x_n; \theta)$$

- If $X_1, \ldots, X_n$ are independent observations, then

$$\hat{\theta} = \arg\max_{\theta} \prod_{i=1}^{n} f_{X_i}(x_i; \theta)$$

Parameter Estimation

- The maximum likelihood (ML) estimate is the parameter value that makes the data most likely

$$\hat{\theta} = \arg\max_{\theta} f_{X_1,\ldots,X_n}(x_1, x_2, \ldots, x_n; \theta)$$

- If $X_1, \ldots, X_n$ are independent observations, then

$$\hat{\theta} = \arg\max_{\theta} \prod_{i=1}^{n} f_{X_i}(x_i; \theta)$$

$$= \arg\max_{\theta} \sum_{i=1}^{n} \log\left(f_{X_i}(x_i; \theta) \right)$$

Parameter Estimation

Suppose that the time it takes for a certain model of hard drive to fail is an exponential random variable

We would like to estimate the unknown parameter λ based on n independent observations $X_1, \ldots, X_n$

What is the maximum likelihood estimate?

(we didn't get to work through the answer in class before time ran out, but we will come back to this…)

Maximum Likelihood is Consistent

Consistency: If θ is the true value of the parameter and θ_n is the maximum likelihood estimate after n observations, then for any $\varepsilon > 0$,

$$\lim_{n \to \infty} P\left(\left|\theta_n - \theta\right| \geq \varepsilon\right) = 0$$

Maximum Likelihood is Consistent

Consistency: If θ is the true value of the parameter and θ_n is the maximum likelihood estimate after n observations, then for any $\varepsilon > 0$,

$$\lim_{n \to \infty} P\left(\left|\theta_n - \theta\right| \geq \varepsilon\right) = 0$$

Translation: As the number of observations gets large, the maximum likelihood estimate gets closer and closer to the true parameter value – clearly desirable for an estimate.