Review of Basic Statistical Concepts

- Statistical Inference
- Point Estimation
- Estimation Error
- Maximum Likelihood Estimate
- Expectation-Maximization (EM)
- Bayes Theorem
- Similarity and Evaluation Measures

Review of Basic Statistical Concepts

- Statistical Inference
- Point Estimation
- Estimation Error
- Maximum Likelihood Estimate
- Expectation-Maximization (EM)
- Bayes Theorem
- Similarity and Evaluation Measures

Statistical Inference

- More fundamental concepts
  - Population
  - Sample

Statistical Inference

- Usually the population is not known completely.

- How to know its parameters?

Statistical Inference

- Usually the population is not known completely.
- We can obtain information about population parameters, by using samples drawn from it.
- Statistical inference deals with such problems.
  - To draw conclusions or inferences about the unknown parameters of the populations from the limited information contained in the sample.

Review of Basic Statistical Concepts

- Statistical Inference
- **Point Estimation**
- Estimation Error
- Maximum Likelihood Estimate
- Expectation-Maximization (EM)
- Bayes Theorem
- Similarity and Evaluation Measures

Estimate

An estimate is a numerical value of the unknown parameter, obtained by applying a formula (estimator) to a particular sample.

If θ is a parameter, $\widehat{\theta}$ denotes its estimate

Estimator

- A rule used to estimate a numerical value is called estimator.

The estimator of mean is given below:

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

E.g., X_i is the height of person i.

Estimate vs. Estimator

Example: Let a sample of size 5 be 2, 4, 5, 9, 10. Then an estimate of the population mean μ, obtained by applying an estimator, is:

$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$ → Estimator

$$\bar{X} = \frac{2 + 4 + 5 + 9 + 10}{5}$$

$\bar{X} = \frac{30}{5} = 6$ → Estimate

Point Estimation Summary

- ***Point Estimate:*** to estimate a population parameter.
- May be made by calculating the parameter for a sample.

Point Estimation Summary

- ***Point Estimate:*** to estimate a population parameter.
- May be made by calculating the parameter for a sample.

- May be used to predict values for the missing data.
- E.g.,
  - A company contains 100 employees
  - 99 have salary information
  - Mean salary of these is $50,000
  - Use $50,000 as value of remaining employee's salary.

Point Estimation Summary

- ***Point Estimate:*** to estimate a population parameter.
- May be made by calculating the parameter for a sample.

- May be used to predict values for the missing data.
- E.g.,
  - A company contains 100 employees
  - 99 have salary information
  - Mean salary of these is $50,000
  - Use $50,000 as value of remaining employee's salary.

Is this a good idea?

Review of Basic Statistical Concepts

- Statistical Inference
- Point Estimation
- **Estimation Error**
- Maximum Likelihood Estimate
- Expectation-Maximization (EM)
- Bayes Theorem
- Similarity and Evaluation Measures

Estimation Error

- ***Bias:*** Difference between expected value and actual value.

$$\boxed{Bias = E(\hat{\Theta}) - \Theta}$$

Estimation Error

- ***Bias:*** Difference between expected value and actual value.

$$\boxed{Bias = E(\hat{\Theta}) - \Theta}$$

- ***Mean Squared Error (MSE):*** expected value of the squared difference between the estimate and the actual value:

$$\boxed{MSE(\hat{\Theta}) = E(\hat{\Theta} - \Theta)^2}$$

Estimation Error

- ***Bias:*** Difference between expected value and actual value.

$$\boxed{Bias = E(\hat{\Theta}) - \Theta}$$

- ***Mean Squared Error (MSE):*** expected value of the squared difference between the estimate and the actual value:

$$\boxed{MSE(\hat{\Theta}) = E(\hat{\Theta} - \Theta)^2}$$

- Why square?

Estimation Error

- ***Bias:*** Difference between expected value and actual value.

$$\boxed{Bias = E(\hat{\Theta}) - \Theta}$$

- ***Mean Squared Error (MSE):*** expected value of the squared difference between the estimate and the actual value:

$$\boxed{MSE(\hat{\Theta}) = E(\hat{\Theta} - \Theta)^2}$$

- Why square?
- Root Mean Square Error (RMSE)

Review of Basic Statistical Concepts

- Statistical Inference
- Point Estimation
- Estimation Error
- Maximum Likelihood Estimate
- Expectation-Maximization (EM)
- Bayes Theorem
- Similarity and Evaluation Measures

Maximum Likelihood Estimate (MLE)

- Obtain parameter estimates that maximize the probability that the sample data occurs for the specific model.

Maximum Likelihood Estimate (MLE)

- Obtain parameter estimates that maximize the probability that the sample data occurs for the specific model.
- Joint probability for observing the sample data by multiplying the individual probabilities.

  Likelihood function:

$$\boxed{L(\Theta \mid x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i \mid \Theta)}$$

- Maximize L.

Maximum Likelihood Estimate (MLE)

- Obtain parameter estimates that maximize the probability that the sample data occurs for the specific model.
- Joint probability for observing the sample data by multiplying the individual probabilities.

Likelihood function:

$$\boxed{L(\Theta \mid x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i \mid \Theta)}$$

- Maximize L.

There is an assumption here. What is it?

MLE Example

- Coin toss five times: {H, H, H, H, T}
- Assuming a perfect coin with H and T equally likely, the likelihood of this sequence is:

$$\boxed{L(p \mid 1,1,1,1,0) = \prod_{i=1}^{5} 0.5 = 0.03.}$$

MLE Example

- Coin toss five times: {H, H, H, H, T}
- Assuming a perfect coin with H and T equally likely, the likelihood of this sequence is:

$$L(p \mid 1,1,1,1,0) = \prod_{i=1}^{5} 0.5 = 0.03.$$

- However if the probability of a H is 0.8 then:

$$L(p \mid 1,1,1,1,0) = 0.8 \times 0.8 \times 0.8 \times 0.8 \times 0.2 = 0.08.$$

MLE Example

- Coin toss five times: {H, H, H, H, T}
- Assuming a perfect coin with H and T equally likely, the likelihood of this sequence is:

$$\boxed{L(p \mid 1,1,1,1,0) = \prod_{i=1}^{5} 0.5 = 0.03.}$$

- However if the probability of a H is 0.8 then:

$$\boxed{L(p \mid 1,1,1,1,0) = 0.8 \times 0.8 \times 0.8 \times 0.8 \times 0.2 = 0.08.}$$

How do we estimate the probability of a H?

MLE Example (cont'd)

- General likelihood formula:

$$L(p \mid x_1, \ldots, x_5) = \prod_{i=1}^{5} p^{x_i} \, (1-p)^{1-x_i} = p^{\sum_{i=1}^{5} x_i} \, (1-p)^{5-\sum_{i=1}^{5} x_i}.$$

MLE Example (cont'd)

- General likelihood formula:

$$L(p \mid x_1, ..., x_5) = \prod_{i=1}^{5} p^{x_i} \, (1-p)^{1-x_i} = p^{\sum_{i=1}^{5} x_i} \, (1-p)^{5-\sum_{i=1}^{5} x_i}.$$

$$l(p) = log L(p) = \sum_{i=1}^{5} x_i log(p) + (5 - \sum_{i=1}^{5} x_i) \; log(1-p)$$

MLE Example (cont'd)

- General likelihood formula:

$$L(p \mid x_1, \ldots, x_5) = \prod_{i=1}^{5} p^{x_i} \, (1-p)^{1-x_i} = p^{\sum_{i=1}^{5} x_i} \, (1-p)^{5-\sum_{i=1}^{5} x_i}.$$

$$l(p) = logL(p) = \sum_{i=1}^{5} x_i log(p) + (5 - \sum_{i=1}^{5} x_i) \; log(1-p)$$

$$\frac{\partial l(p)}{\partial p} = \sum_{i=1}^{5} \frac{x_i}{p} - \frac{5 - \sum_{i=1}^{5} x_i}{1-p}.$$

MLE Example (cont'd)

- General likelihood formula:

$$L(p \mid x_1, ..., x_5) = \prod_{i=1}^{5} p^{x_i} \, (1-p)^{1-x_i} = p^{\sum_{i=1}^{5} x_i} \, (1-p)^{5-\sum_{i=1}^{5} x_i}.$$

$$l(p) = logL(p) = \sum_{i=1}^{5} x_i log(p) + (5 - \sum_{i=1}^{5} x_i) \; log(1-p)$$

$$\frac{\partial l(p)}{\partial p} = \sum_{i=1}^{5} \frac{x_i}{p} - \frac{5 - \sum_{i=1}^{5} x_i}{1-p}.$$

$$p = \frac{\sum_{i=1}^{5} x_i}{5}$$

- MLE Estimate for p is then 4/5 = 0.8

Review of Basic Statistical Concepts

- Statistical Inference
- Point Estimation
- Estimation Error
- Maximum Likelihood Estimate
- Expectation-Maximization (EM)
- Bayes Theorem
- Similarity and Evaluation Measures

Expectation-Maximization (EM)

- Solves estimation with incomplete data.

- Key Idea:
  - Obtain initial estimates for parameters.
  - Iteratively use estimates for missing data and continue until convergence.

EM Example

$\{1, 5, 10, 4\}$; $n = 6$ $k = 4$; **Guess** $\hat{\mu}^0 = 3$.

$$\hat{\mu}^1 = \frac{\sum_{i=1}^{k} x_i}{n} + \frac{\sum_{i=k+1}^{n} x_i}{n} = 3.33 + \frac{3+3}{6} = 4.33$$

$$\hat{\mu}^2 = \frac{\sum_{i=1}^{k} x_i}{n} + \frac{\sum_{i=k+1}^{n} x_i}{n} = 3.33 + \frac{4.33+4.33}{6} = 4.77$$

$$\hat{\mu}^3 = \frac{\sum_{i=1}^{k} x_i}{n} + \frac{\sum_{i=k+1}^{n} x_i}{n} = 3.33 + \frac{4.77+4.77}{6} = 4.92$$

$$\hat{\mu}^4 = \frac{\sum_{i=1}^{k} x_i}{n} + \frac{\sum_{i=k+1}^{n} x_i}{n} = 3.33 + \frac{4.92+4.92}{6} = 4.97$$

EM Algorithm

Input:

$\Theta = \{\theta_1, ..., \theta_p\}$ **//Parameters to be Estimated**

$X_{obs} = \{x_1, ..., x_k\}$ **//Input Database Values Observed**

$X_{miss} = \{x_{k+1}, ..., x_n\}$ **//Input Database Values Missing**

Output:

$\hat{\Theta}$ **//Estimates for Θ**

EM Algorithm:

```
i := 0;
Obtain initial parameter MLE estimate, Θ̂^i;
repeat
    Estimate missing data, X̂^i_miss;
    i++;
    Obtain next parameter estimate, θ̂^i to maximize data;
until estimate converges;
```

Review of Basic Statistical Concepts

- Statistical Inference
- Point Estimation
- Estimation Error
- Maximum Likelihood Estimate
- Expectation-Maximization (EM)
- Bayes Theorem
- Similarity and Evaluation Measures

Bayes Theorem Example

- Credit authorizations (hypotheses): h_1=authorize purchase, h_2 = authorize after further identification, h_3=do not authorize, h_4= do not authorize but contact police
- Task: Assign a label for each combination of credit (col.) and income (row):

	1	2	3	4
Excellent	x_1	x_2	x_3	x_4
Good	x_5	x_6	x_7	x_8
Bad	x_9	x_{10}	x_{11}	x_{12}

Bayes Example(cont'd)

- Training Data:

ID	Income	Credit	Class	x_i
1	4	Excellent	h_1	x_4
2	3	Good	h_1	x_7
3	2	Excellent	h_1	x_2
4	3	Good	h_1	x_7
5	4	Good	h_1	x_8
6	2	Excellent	h_1	x_2
7	3	Bad	h_2	x_{11}
8	2	Bad	h_2	x_{10}
9	3	Bad	h_3	x_{11}
10	1	Bad	h_4	x_9

From training data:

$P(h_1)$ = ?; $P(h_2)$=?; $P(h_3)$=?; $P(h_4)$=?.

Bayes Example(cont'd)

- Training Data:

ID	Income	Credit	Class	x_i
1	4	Excellent	h_1	x_4
2	3	Good	h_1	x_7
3	2	Excellent	h_1	x_2
4	3	Good	h_1	x_7
5	4	Good	h_1	x_8
6	2	Excellent	h_1	x_2
7	3	Bad	h_2	x_{11}
8	2	Bad	h_2	x_{10}
9	3	Bad	h_3	x_{11}
10	1	Bad	h_4	x_9

From training data:

$P(h_1) = 60\%$; $P(h_2)=20\%$; $P(h_3)=10\%$; $P(h_4)=10\%$.

Bayes Example(cont'd)

- How to predict the class for X_4?

ID	Income	Credit	Class	x_i
1	4	Excellent	h_1	x_4
2	3	Good	h_1	x_7
3	2	Excellent	h_1	x_2
4	3	Good	h_1	x_7
5	4	Good	h_1	x_8
6	2	Excellent	h_1	x_2
7	3	Bad	h_2	x_{11}
8	2	Bad	h_2	x_{10}
9	3	Bad	h_3	x_{11}
10	1	Bad	h_4	x_9

Bayes Example(cont'd)

- How to predict the class for X_4?
  - Calculate $P(h_i | X_4)$ for all h_i.
  - Place X_4 in class with largest value.

ID	Income	Credit	Class	x_i
1	4	Excellent	h_1	x_4
2	3	Good	h_1	x_7
3	2	Excellent	h_1	x_2
4	3	Good	h_1	x_7
5	4	Good	h_1	x_8
6	2	Excellent	h_1	x_2
7	3	Bad	h_2	x_{11}
8	2	Bad	h_2	x_{10}
9	3	Bad	h_3	x_{11}
10	1	Bad	h_4	x_9

Bayes Example(cont'd)

- How to predict the class for X_4?
  - Calculate $P(h_i | X_4)$ for all h_i.
  - Place X_4 in class with largest value.
  - In Math:
    - $P(h_1 | x_4)=(P(x_4 | h_1)(P(h_1))/P(x_4)$ $=(1/6)(0.6)/0.1=1.$
    - x_4 in class h_1.

ID	Income	Credit	Class	x_i
1	4	Excellent	h_1	x_4
2	3	Good	h_1	x_7
3	2	Excellent	h_1	x_2
4	3	Good	h_1	x_7
5	4	Good	h_1	x_8
6	2	Excellent	h_1	x_2
7	3	Bad	h_2	x_{11}
8	2	Bad	h_2	x_{10}
9	3	Bad	h_3	x_{11}
10	1	Bad	h_4	x_9

Bayes Example(cont'd)

- How to predict the class for X_4?
  - Calculate $P(h_i | X_4)$ for all h_i.
  - Place X_4 in class with largest value.
  - In Math:
    - $\boxed{P(h_1 | x_4)=(P(x_4 | h_1)(P(h_1))/P(x_4)}$

      $=(1/6)(0.6)/0.1=1.$
    - x_4 in class h_1.

ID	Income	Credit	Class	x_i
1	4	Excellent	h_1	x_4
2	3	Good	h_1	x_7
3	2	Excellent	h_1	x_2
4	3	Good	h_1	x_7
5	4	Good	h_1	x_8
6	2	Excellent	h_1	x_2
7	3	Bad	h_2	x_{11}
8	2	Bad	h_2	x_{10}
9	3	Bad	h_3	x_{11}
10	1	Bad	h_4	x_9

Bayes Theorem

Review of Basic Statistical Concepts

- Statistical Inference
- Point Estimation
- Estimation Error
- Maximum Likelihood Estimate
- Expectation-Maximization (EM)
- Bayes Theorem
- Similarity and Evaluation Measures

Similarity Measures

- Determine similarity between two objects.
- Similarity characteristics:

- $\forall t_i \in D, sim(t_i, t_i) = 1$
- $\forall t_i, t_j \in D, sim(t_i, t_j) = 0$ **if** t_i **and** t_j **are not alike at all.**
- $\forall t_i, t_j, t_k \in D, sim(t_i, t_j) < sim(t_i, t_k)$ **if** t_i **is more like** t_k **than it is like** t_j**.**

- Alternatively, distance measure measures how unlike or dissimilar objects are.

Similarity Measures

Dice: $$sim(t_i, t_j) = \frac{2\sum_{h=1}^{k} t_{ih}t_{jh}}{\sum_{h=1}^{k} t_{ih}^2 + \sum_{h=1}^{k} t_{jh}^2}$$

Jaccard: $$sim(t_i, t_j) = \frac{\sum_{h=1}^{k} t_{ih}t_{jh}}{\sum_{h=1}^{k} t_{ih}^2 + \sum_{h=1}^{k} t_{jh}^2 - \sum_{h=1}^{k} t_{ih}t_{jh}}$$

Cosine: $$sim(t_i, t_j) = \frac{\sum_{h=1}^{k} t_{ih}t_{jh}}{\sqrt{\sum_{h=1}^{k} t_{ih}^2 \sum_{h=1}^{k} t_{jh}^2}}$$

Overlap: $$sim(t_i, t_j) = \frac{\sum_{h=1}^{k} t_{ih}t_{jh}}{min(\sum_{h=1}^{k} t_{ih}^2, \sum_{h=1}^{k} t_{jh}^2)}$$

Distance Measures

- Measure dissimilarity between objects

$$\textbf{Euclidean:}\ \ dis(t_i, t_j) = \sqrt{\sum_{h=1}^{k} (t_{ih} - t_{jh})^2}$$

$$\textbf{Manhattan:}\ \ dis(t_i, t_j) = \sum_{h=1}^{k} \mid (t_{ih} - t_{jh}) \mid$$

Distance Measures

- Measure dissimilarity between objects

$$\textbf{Euclidean:}\ \ dis(t_i, t_j) = \sqrt{\sum_{h=1}^{k} (t_{ih} - t_{jh})^2}$$

$$\textbf{Manhattan:}\ \ dis(t_i, t_j) = \sum_{h=1}^{k} \mid (t_{ih} - t_{jh}) \mid$$

Why is it called Manhattan distance?

References

- Previous CSE 5243 course offered by Prof. Srinivasan Parthasarathy @OSU:

http://web.cse.ohio-state.edu/~parthasarathy.2/674/

- Point Estimation on SlidesShare:

https://www.slideshare.net/ShahabYaseen/point-estimation-48241348