SMAI-S25-06: Data and Distribution

C. V. Jawahar

IIIT Hyderabad

January 24, 2025

Recap:

- Two problems of interest:
  - Learn a function $y = f(\mathbf{W}, \mathbf{x})$ from the data.
  - Learn Feature Transformation as a step to find useful representations.
- Three Classification Schemes:
  - Nearest Neighbour Algorithm
  - Linear Classification
  - Decide as ω_1 if $P(\omega_1|\mathbf{x}) \geq P(\omega_2|\mathbf{x})$ else ω_2. More today
- Performance Metrics:
  - Classification: Accuracy, TP/FP etc., Confusion Matrix; Ranking: Precision, Recall, F-Score, AP
- Supervised Learning:
  - Notion of Training, Validation and Testing
  - Notion of Loss Function (soon)
  - Role of Optimization (soon)

Probability: Recap

- Probability of an Event
- Discrte and Continous Distributions (Normal, Bernoulli)
- Joint and Conditional probabilities
- Sum and Product Rule
- Total Probability
- Bayes Theorem

Normal Distribution

Univariate Normal Distribution $\mathcal{N}(\mu, \sigma)$

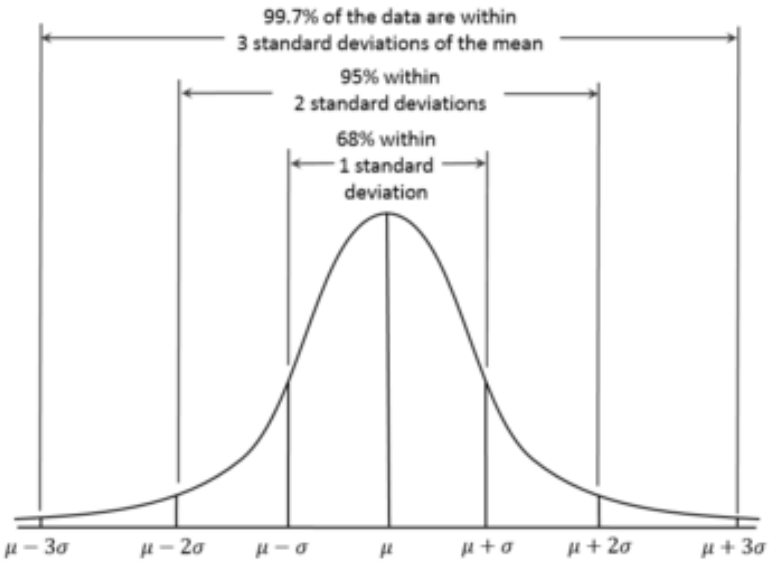
$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

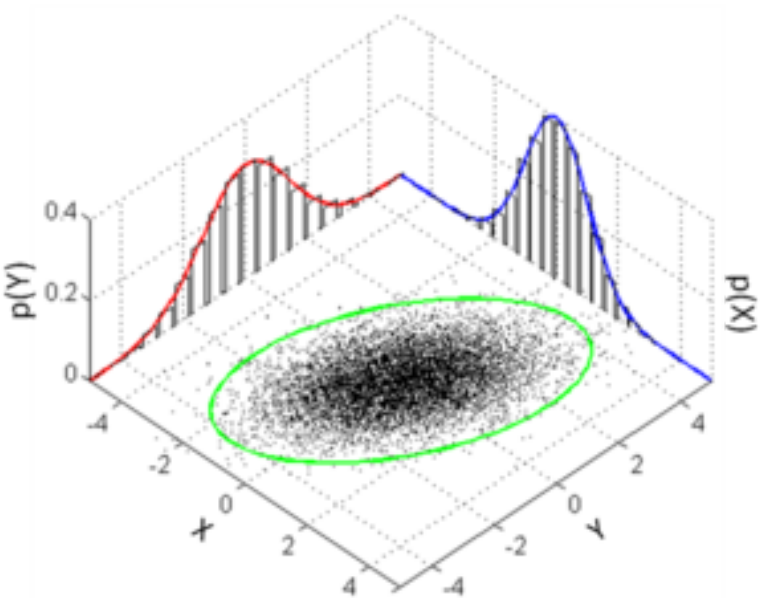
Multivariate Normal Distribution $\mathcal{N}(\mu, \Sigma)$ with $\mathbf{x} \in R^d$

$$p(\mathbf{x}|\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}([\mathbf{x}-\mu]^T \Sigma^{-1} [\mathbf{x}-\mu])}$$

- **Appreciate Mean**; Interpretation? Properties? Dimensionality
- **Appreciate Variance/Covariance** Interpretation? Properties? Dimensionality

Examples


99.7% of the data are within
3 standard deviations of the mean
95% within
2 standard deviations
68% within
1 standard
deviation
$\mu - 3\sigma$
$\mu - 2\sigma$
$\mu - \sigma$
μ
$\mu + \sigma$
$\mu + 2\sigma$
$\mu + 3\sigma$


p(Y)
0.4
0.2
0
-4
-2
0
2
4
X
p(X)
4
2
0
-2
-4
Y

Bayesian Decision

Bayes Rule:

Bayes rule states that the joint probability of $\mathbf{x}$ and ω_i, denoted as $p(\mathbf{x}, \omega_i)$ is given by:

$$p(\mathbf{x}, \omega_i) = p(\mathbf{x}/\omega_i).P(\omega_i) = P(\omega_i/\mathbf{x}).P(\mathbf{x})$$

We can rewrite the second equality as:

$$P(\omega_i/\mathbf{x}) = \frac{p(\mathbf{x}/\omega_i).P(\omega_i)}{P(\mathbf{x})}$$

Here the L.H.S is the posterior probability of class ω_i after observing $\mathbf{x}$. Bayes decision rule says to choose that ω_i which maximises the posterior probability. The above equation may also be written as:

$$P(\omega_i/\mathbf{x}) = \frac{p(\mathbf{x}/\omega_i).P(\omega_i)}{\Sigma_{j=1}^{c} p(\mathbf{x}/\omega_j).P(\omega_j)}$$

Technically

- The simple (or Naive) Decision Rule:
  - Decide ω_1 is $P(\omega_1) > P(\omega_2)$
  - Decide ω_2 is $P(\omega_2) \geq P(\omega_1)$
- This is a near-blind decision making. Only the prior probabilities are used. Evidences are discarded.
- In most cases, we will have the measurement or feature vector to aid the classification.
- Bayes decision rule says:
  Decide ω_1 if $P(\omega_1|\mathbf{x}) > P(\omega_2|\mathbf{x})$; Otherwise decide ω_2

Bayesian Decision Making

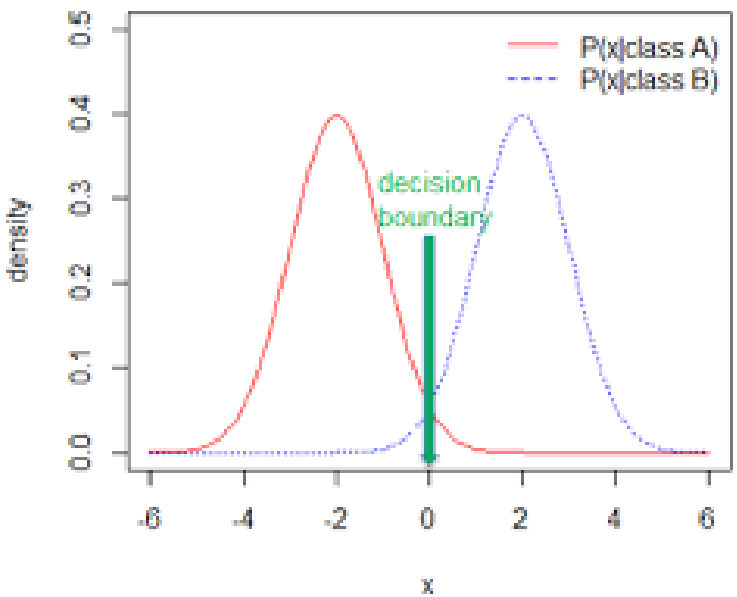
Decide as ω_1 if

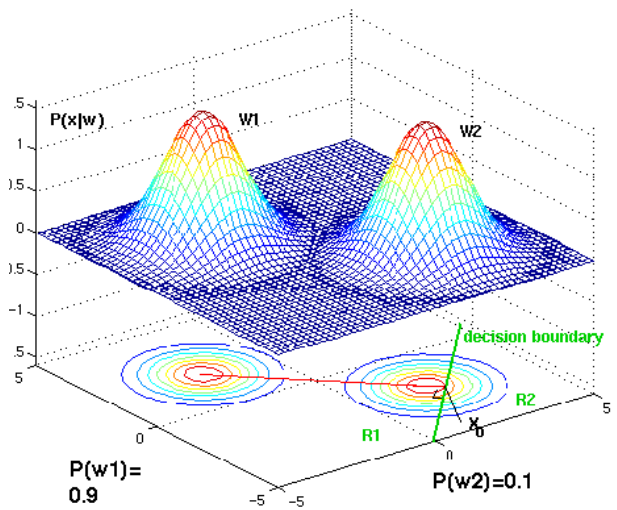
$$P(\omega_1|\mathbf{x}) \geq P(\omega_2|\mathbf{x})$$

Else ω_2

Bayesian Decision is "**Optimal** Minimal Erroe Classification". Why?

Optimal Bayesian Classifier





Simplifying Bayesian Decision Making

Decide as ω_1 if $P(\omega_1|\mathbf{x}) \geq P(\omega|\mathbf{x})$ Else ω_2

Consider a problem with class 1 and 2 means as μ_1 and μ_2. Assume they have equal variance $\sigma_1 = \sigma_2 = \sigma$. What will be the decision rule?

Decide as ω_1 if

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma^2}} . \geq \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu_2)^2}{2\sigma^2}}$$

How do we rewrite this decision as a "Nearest Neighbour" rule or like a "Linear Classifier" Decision?

Note: MAP and ML Classification

Maximium A Posteriori (MAP) and Maximum Likelihood (ML) are two popular formulations for estimating parameters (here decisions).

MAP: Decide as ω_1 if

$$P(\omega_1/\mathbf{x}) \geq P(\omega_2/\mathbf{x})$$

$$\frac{p(\mathbf{x}/\omega_1).P(\omega_1)}{P(\mathbf{x})} \geq \frac{p(\mathbf{x}/\omega_2).P(\omega_2)}{P(\mathbf{x})}$$

$$p(\mathbf{x}/\omega_1).P(\omega_1) \geq p(\mathbf{x}/\omega_2).P(\omega_2)$$

When Prior Probabailites are same i.e., $P(\omega_1) = P(\omega_2)$

Maximum Liklihood

$$p(\mathbf{x}/\omega_1) \geq p(\mathbf{x}/\omega_2)$$

Problem 1

Over years, we have figured out HYD temperature in Jan and May are $\mathcal{N}(23, \sigma^2)$ and $\mathcal{N}(33, \sigma^2)$ (i.e, Normal, mean 23 and 33; Variance the same).

Q: We have 100 days of data from Jan and 100 days from May, but not labelled. We want a classifier as:

"If temp $< \theta$, then Jan else May"

What should be the value of θ intuitively?

- 28 $(= \frac{23+33}{2})$
- Less than 28.
- More than 28.

Why?

Problem 2

A disease occurs with a probability of 0.4 (i.e., it is present in 40% of the population). You have a test that detects the disease with a probability 0.6, and produces a false positive with probability of 0.1. What is the (posterior) probability that the test comes back positive.

Hint: S is the event that you are sick; P is the event that test comespositive.

$$P(S|P) = \frac{P(P|S)P(S)}{P(P)} = \frac{P(P|S)P(S)}{P(P|S)P(S) + P(P|\bar{S})P(\bar{S})}$$

(a) 0.6 (b) 0.7 (c) 0.8 (d) 0.9 (e) 0.95

Problem 3

You are planning a picnic today, but the morning is cloudy

- 50% of all rainy days start off cloudy.
- But cloudy mornings are common (about 40% of days start cloudy)
- This is usually a dry month (only 3 of 30 days tend to be rainy, or 10%)

What is the chance of rain during the day?

(a) 10% (b) 12.5% (c) 15% (d) $> 20\%$ (e) $< 20\%$

Hint: Bayes Rule:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Problem 4

In a TV Game show, a contestant selects one of three doors; behind one of the doors there is a prize,and behind the other two there are no prizes. After the contestant selects a door, the game-show host opens one of the remaining doors, and reveals that there is no prize behind it. The host then asks the contestant whether the you want to SWITCH their choice to the other unopened door, or STICK to their original choice. [1]

Use Bayes Theorem: https://blogs.cornell.edu/info2040/2022/11/10/the-monty-hall-problem-using-bayes-theorem/

What should we advise? What is the prob. of win if the candidate switch:

- (a) $\frac{1}{3}$ since all the doors are equally likely. Don't switch
- (b) $\frac{1}{2}$ since there are only two left, both are equally likely, no advantage in switching.
- (c) $\frac{2}{3}$. Prefer switching. Bayes says so.
- (d) $\frac{1}{3}$. Don't switching. Bayes says so.
- (e) None of the above.

[1] A very popular problem on internet from khan academy to mit lecture notes!. Appreciate the role of evidence, specially if the answer is not intuitive.

Problem 5: Streaming Data

Consider a situation when we continue to get one sample at a time. We call such situations as "streaming data".
We have mean (μ_N) and variance (σ^2_N) computed and available at sample N.

Now we get the $N+1$ sample. How do we compute the new mean? Ans:

$$Ans: \mu_{N+1} = \frac{\mu_N \times N + x_{N+1}}{N+1}$$

How do we compute σ^2_{N+1}?
Where do we need such "online" computations?

Questions? Comments?