CSCI 6521
Advanced Machine Learning I

Chapter 01:

Generative Model

Md Tamjidul Hoque

Syllabus

Textbook for optional reading:

1. Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd Edition., by Aurélien Géron, 2023, book@UNO link.
2. Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play, 2nd Edition, by David Foster.
3. Practical Deep Learning for Cloud, Mobile, and Edge: Real-World AI & Computer-Vision Projects Using Python, Keras & TensorFlow, by Anirudh Koul, Meher Kasam, and Siddha Ganju.
4. The Elements of Statistical Learning, 2nd edn, by Trevor Hastie, Robert Tibshirani and Jerome Friedman. Springer, 2009, online: http://www-stat.stanford.edu/~tibs/ElemStatLearn/
5. Hands On Unsupervised Learning Using Python by Ankur A. Patel.

Mark distribution:

Programming Assignments (3)	→	39% [Any language is acceptable]
Homework Assignments (1)	→	11%
Class Test (3)	→	20% (Best 2 counts)
Final Examination	→	25% (Must attend to pass)
Attendance	→	5%,

Exams: Class tests: (1) Feb/13, (2) Mar/19, (3) Apr/23 [Tentative]
Final Exam: May/02/2024 (Thursday), 5:30 PM – 7:30 PM.

Attendance

- Your attendance at class is required and essential for you to meet course requirements.

- 5% make is allocated for attendance and

- distributed as %5: [90-100%], 4%: [85-90), 3%: [80-85), 2%: [75-80), 1%: [70-75), 0%: <70.

Office Hours

- You can see me after the class:
- **Tuesday, Thursday**, 6:15 PM to 8:15 PM.
- **Wednesday,** 11:00 AM to 1:00 PM.

- Otherwise, feel free to email/phone me
  - *Email: thoque@uno.edu*
  - *Phone:* (504)-280-2406

Assignment

- **The submitted assignment must be your own work.**
  - If you find a problem with Canvas for assignment submission, please only then email me the assignment (*Email: thoque@uno.edu)*
  - You must submit both Hard & Soft copy – if indicated.
- **Bonus**
  - A student who will be able to produce any publishable work (approved based on superior results, recognized by the instructor during the period of this course) related to any given assignment(s) or the topics covered in the class, will be given **10% bonus** marks.
  - Easy venue for submission:
    - InnovateUNO: Oct/Nov conference @UNO.
    - LBRN: April conference, https://lbrn.lsu.edu/
  - For samples, see: https://cs.uno.edu/~tamjid/publications.html

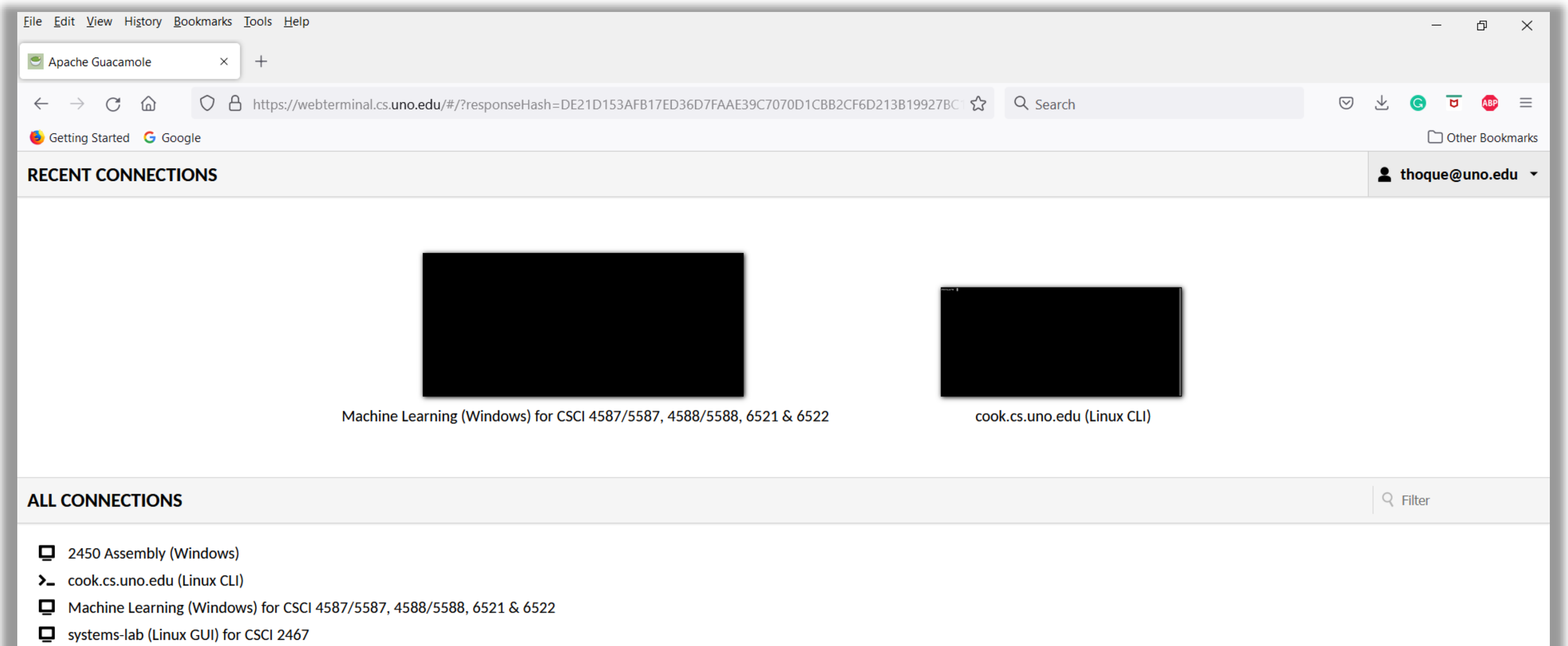
… Assignment

- **Due Dates**
  - Every assignment handed out will be clearly marked with a due date. You are responsible for handing in your assignment on time.

- Late submissions will be assessed at the following rates:
  - 80% for 1-48 hours (2 days max) late,
  - 60% for 49-96 hours (2 days max) late,
  - 40% for 97-144 (2 days max) hours late,
  - 20% for 145-168 hours (1-day max) late.
  - Assignments that are <u>more than a week late</u> will receive no credit.

Course Resources

- Check Moodle for resources
- Please remember:
  - I often update the lecture notes just before teaching in the class – so if you want to save/print the final copy – please do so after covering it in the class.
- Accessing lab computers: webterminal.cs.uno.edu/



Other Offerings

You may be interested in:

- Undergrade Machine Learning (ML) and AI Concentration, click-here for the details.
- For the Graduate Certificate in ML & AI, click-here for the details.
- CSCI 4587/5587 ML I, click-here for a sample syllabus.
- CSCI 4588/5588 ML II, click-here for a sample syllabus.
- CSCI 6522 Advanced ML II, click-here for a sample syllabus.

Topic Covered

- **Introduction to various programming aspects, tools, and platforms of machine learning** (see more here)
- **Regression, Classification, and Optimization**
- **Generative Modeling**
- **Autoencoders**
- **Representation Learning and Generative Learning Using Autoencoders and GANs**
- **Paint**
- **Write – Natural Language Generator**
- **Compose**
- **Advanced Generative Modeling**
- **Feature Detection Using Deep Belief Networks**
- **Time Series Clustering**
- **Real-Time Object Classification on iOS with Core ML**

Generative Model

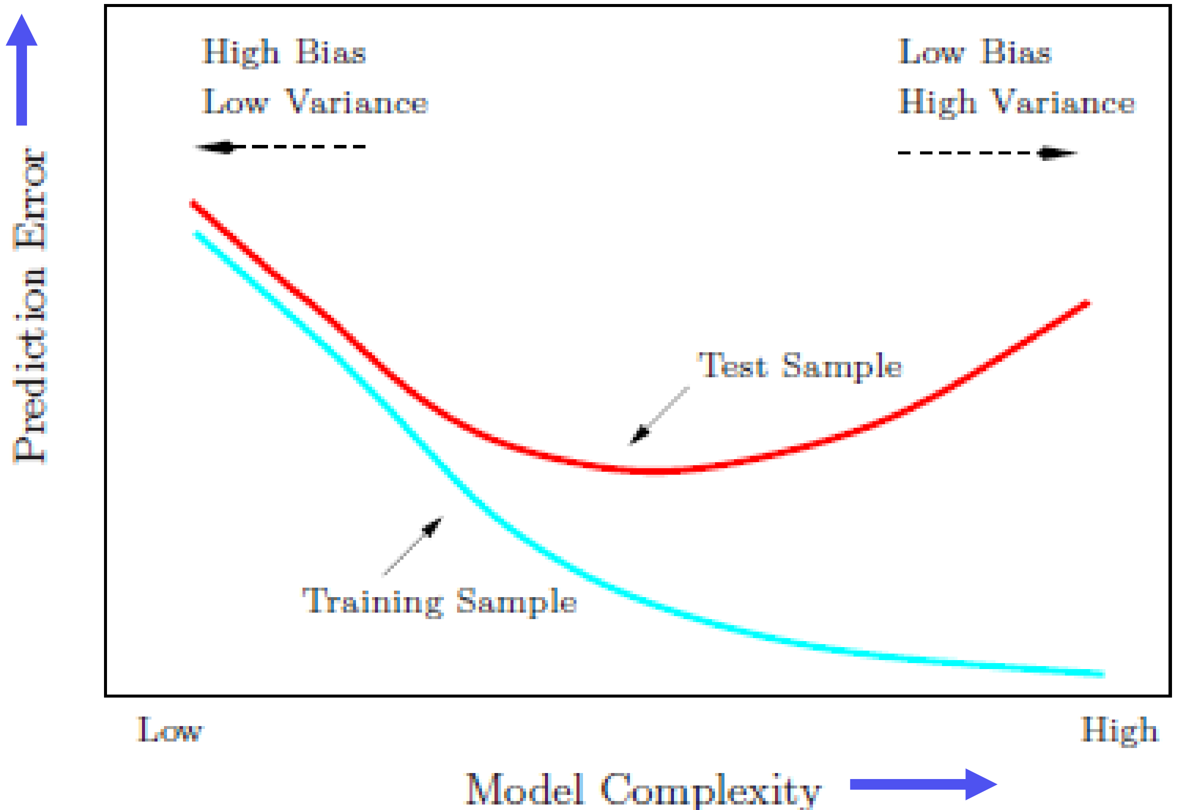
- Some Preliminaries
  - Pages: 1 to 7 from the lecture note.

Bias-Variance Decomposition

$$MSE = Var + bias^2$$

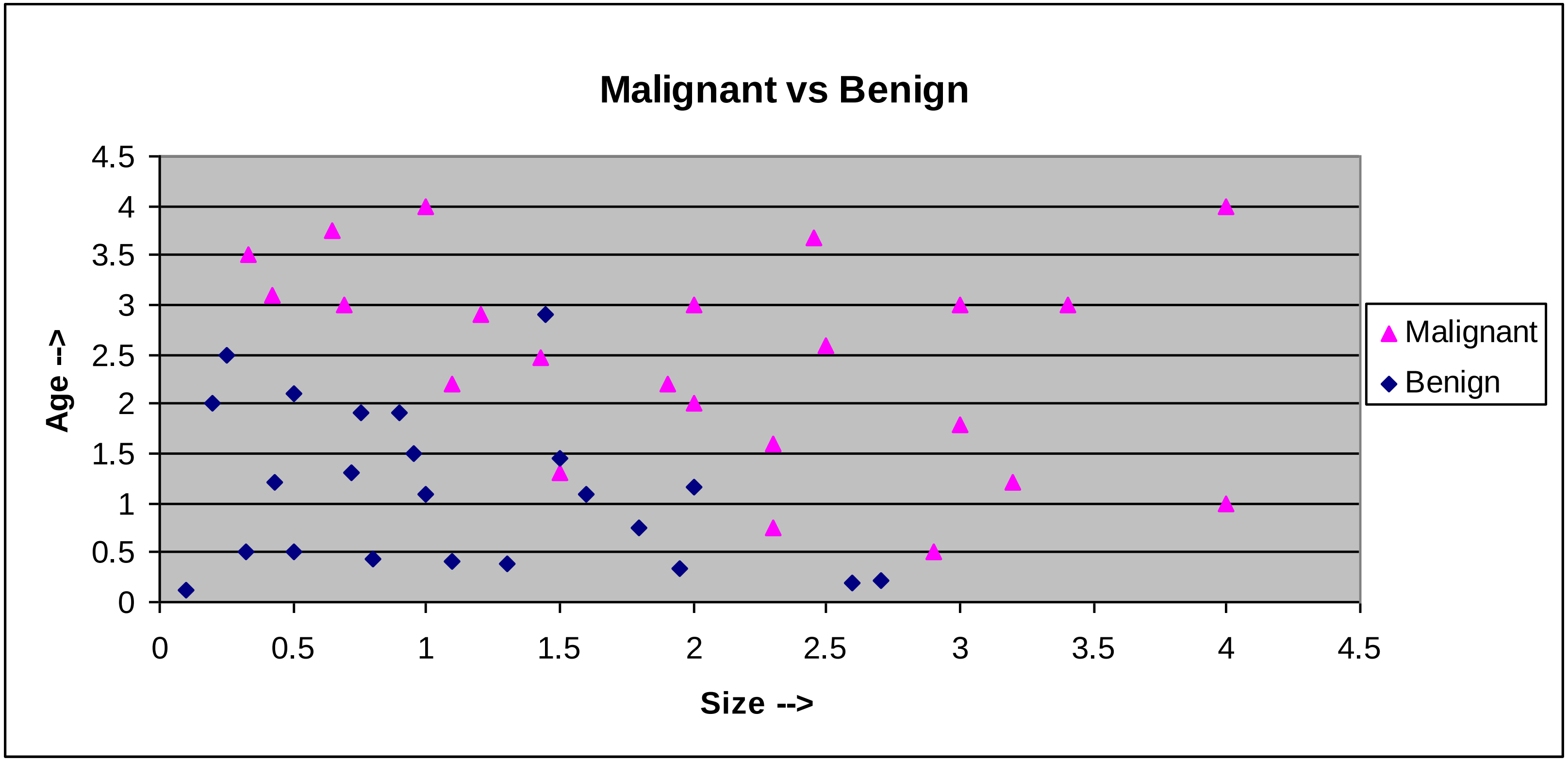
- The variance measures the extent to which the solutions for individual data sets vary around their average and hence this measures the extent to which the function *f*(*X*) is sensitive to the particular choice of data set.

- The bias represents the extent to which the average prediction over all data sets differs from the desired regression function.



Exercise using Weka

- Check *Exercise.zip* in Moodle



- Weka is a java based software to explore machine-learning and Data-mining approaches: http://www.cs.waikato.ac.nz/ml/weka/

Bayes Law/Theorem

$$P(x \mid y) = \frac{P(y \mid x)P(x)}{\sum_{x \in X} P(y \mid x)P(x)}$$

- Example: Suppose, we have invented an influenza tester. Now, say 5% of the population is sneezing due to the cold season. Running test on sneezing people, 90% test returned positive for influenza testing. Given a person is NOT sneezing the test comes out positive 15% of the time. For a positive detection what is the chance that the person is sneezing?

Some Important Preliminaries

- Pages: ~09 to 14 - we will discuss the followings:

  - Gaussian or Normal Distribution
  - Probability Density Function
  - Likelihood Functions
  - Maximum Likelihood Estimation (MLE)

Gaussian Distributions

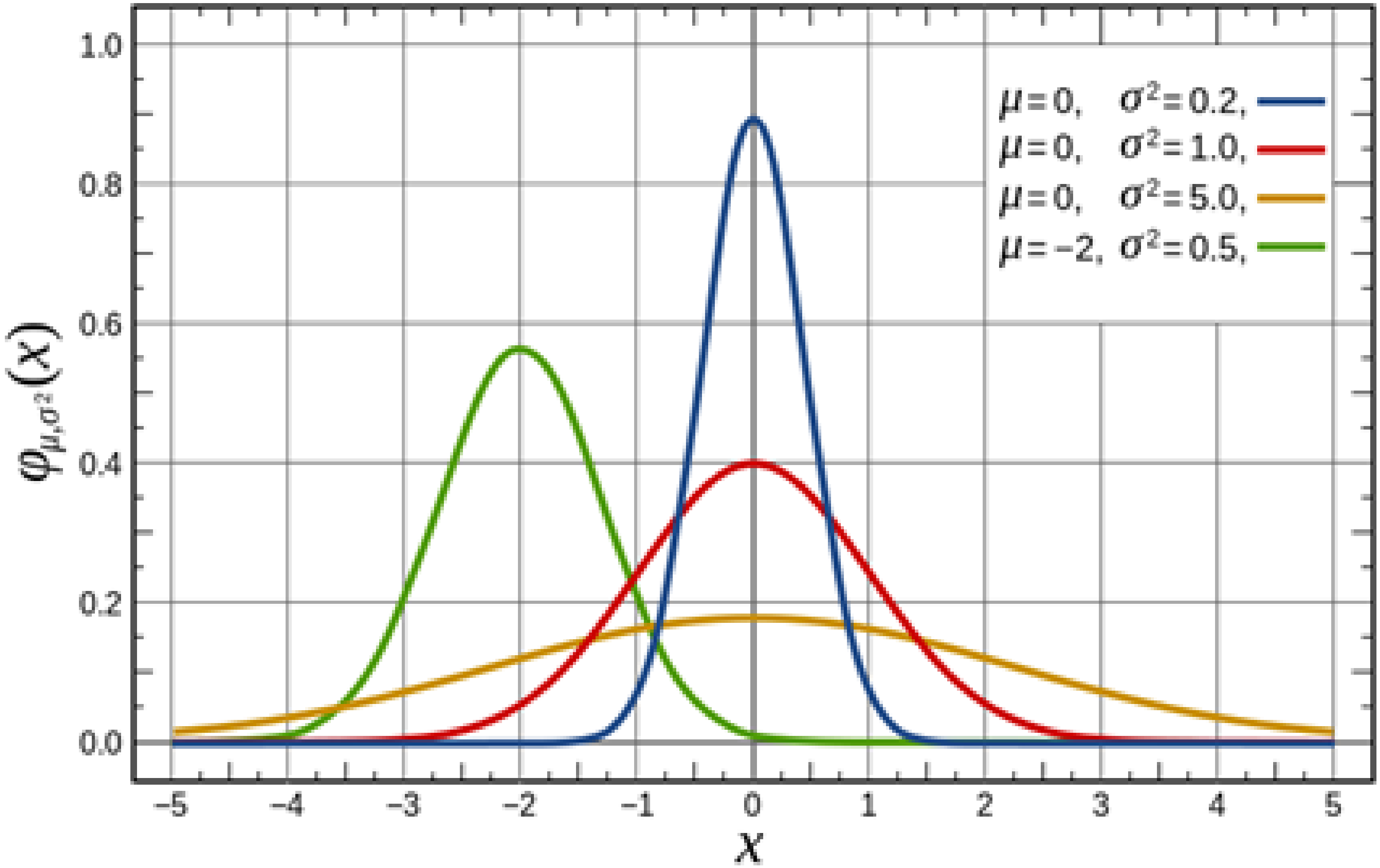


Fig: Different distribution for different values of μ and σ^2 (*ref: wiki*)

Generative Model

- Definition: A generative model describes how a dataset is generated in terms of a probabilistic model. By sampling from this model, we are able to generate new data.

- Example:
  - Suppose we have a dataset containing images of horses.
  - We may wish to build a model that can generate a new image of a horse that has never existed but still looks real because the model has learned the general rules that govern the appearance of a horse.
  - This is the kind of problem that can be solved using generative modeling.

… Generative Model

There has been increased media attention on generative modeling projects such as:

- StyleGAN (GAN stands for Generative Adversarial Networks) from NVIDIA, which is able to create hyper-realistic images of human faces, see https://nvlabs.github.io/stylegan2/versions.html

- the GPT-2 language model from OpenAI, which is able to complete a passage of text given a short introductory paragraph. See “Language Models Are Unsupervised Multitask Learners”, paper here: https://paperswithcode.com/paper/language-models-are-unsupervised-multitask

Generative Model Framework

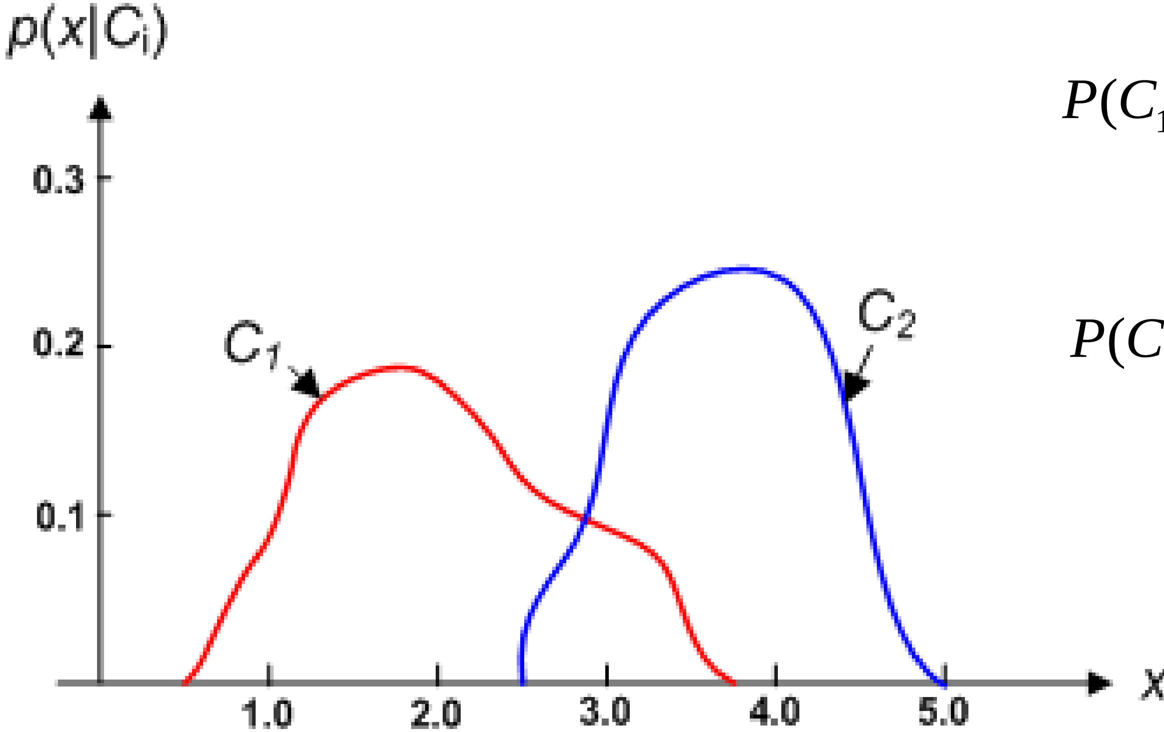
We have a dataset of observations **X**.

- We assume that the observations have been generated according to some unknown distribution, p_{data}.
- A generative model p_{model} tries to mimic p_{data}. If we achieve this goal, we can sample from p_{model} to generate observations that appear to have been drawn from p_{data}.
- We are impressed by p_{model} if:
  - Rule 1: It can generate examples that appear to have been drawn from p_{data}.
  - Rule 2: It can generate examples that are appropriately different from the observations in **X**. In other words, the model shouldn't simply reproduce things it has already seen.

Presentation Overview

- We will discuss the following items mostly by ‘chalk and talk’:
  - Bayes Theorem (review)
  - Naïve Bayes Classifier
  - Laplace Smoothing / Additive smoothing
  - Gaussian Densities (specially, Multivariate)
  - Bernoulli Distribution
  - Gaussian Discriminant Analysis (GDA)

Bayes Law



$$P(C_1 \mid x) = \frac{p(x \mid C_1)P(C_1)}{p(x)}$$

$$P(C_2 \mid x) = \frac{p(x \mid C_2)P(C_2)}{p(x)}$$

Figure 1: Hypothetical class-conditional probability density functions show the probability density of measuring a particular feature value *x* given the pattern is in category C_i. If *x* represent the tumor-*size* (or the *age*) the difference in tumor size (or *age*) of populations of two types of tumor. Density functions are normalized, and thus the area under each curve is 1.0

Naïve Bayes Classifier

Assume, given a set of variables, $X = \{x_1, x_2, ..., x_d\}$, and class sets $C_i = \{C_1, C_{2,} ..., C_k\}$.

$$P(C_i \mid x_1, x_2, ..., x_d) \equiv p(x_1, x_2, ..., x_d \mid C_i) P(C_i)$$

Naïve Bayes assumes that the conditional probabilities of the variables are statistically independent. The likelihood is computed as products:

$$P(X \mid C_i) \equiv \prod_{j=1}^{d} p(x_j \mid C_i)$$

$$P(C_i \mid X) \equiv P(C_i) \prod_{j=1}^{d} p(x_j \mid C_i)$$

$$C_i = \arg\max_{c_i} P(c_i) \prod_{j=1}^{d} p(x_j \mid c_i)$$

Example 1

Table 1: The weather data (in nominal form) is given to decide whether to play outside or not.

outlook	temperature	humidity	Windy	play
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
overcast	hot	high	FALSE	yes
rainy	mild	high	FALSE	yes
rainy	cool	normal	FALSE	yes
rainy	cool	normal	TRUE	no
overcast	cool	normal	TRUE	yes
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
rainy	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

Table 2: Likelihood computation

outlook			temperature			humidity			Windy		
play →	yes	no	play →	yes	no	play →	yes	no	play →	yes	no
sunny	$\frac{2}{9}$	$\frac{3}{5}$	hot	$\frac{2}{9}$	$\frac{2}{5}$	high	$\frac{3}{9}$	$\frac{4}{5}$	TRUE	$\frac{3}{9}$	$\frac{3}{5}$
overcast	$\frac{4}{9}$	$\frac{0}{5}$	mild	$\frac{4}{9}$	$\frac{2}{5}$	normal	$\frac{6}{9}$	$\frac{1}{5}$	FALSE	$\frac{6}{9}$	$\frac{2}{5}$
Rainy	$\frac{3}{9}$	$\frac{2}{5}$	cool	$\frac{3}{9}$	$\frac{1}{5}$						

Normal/Gaussian Distribution

$$N(x \mid \mu, \sigma^2) = p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

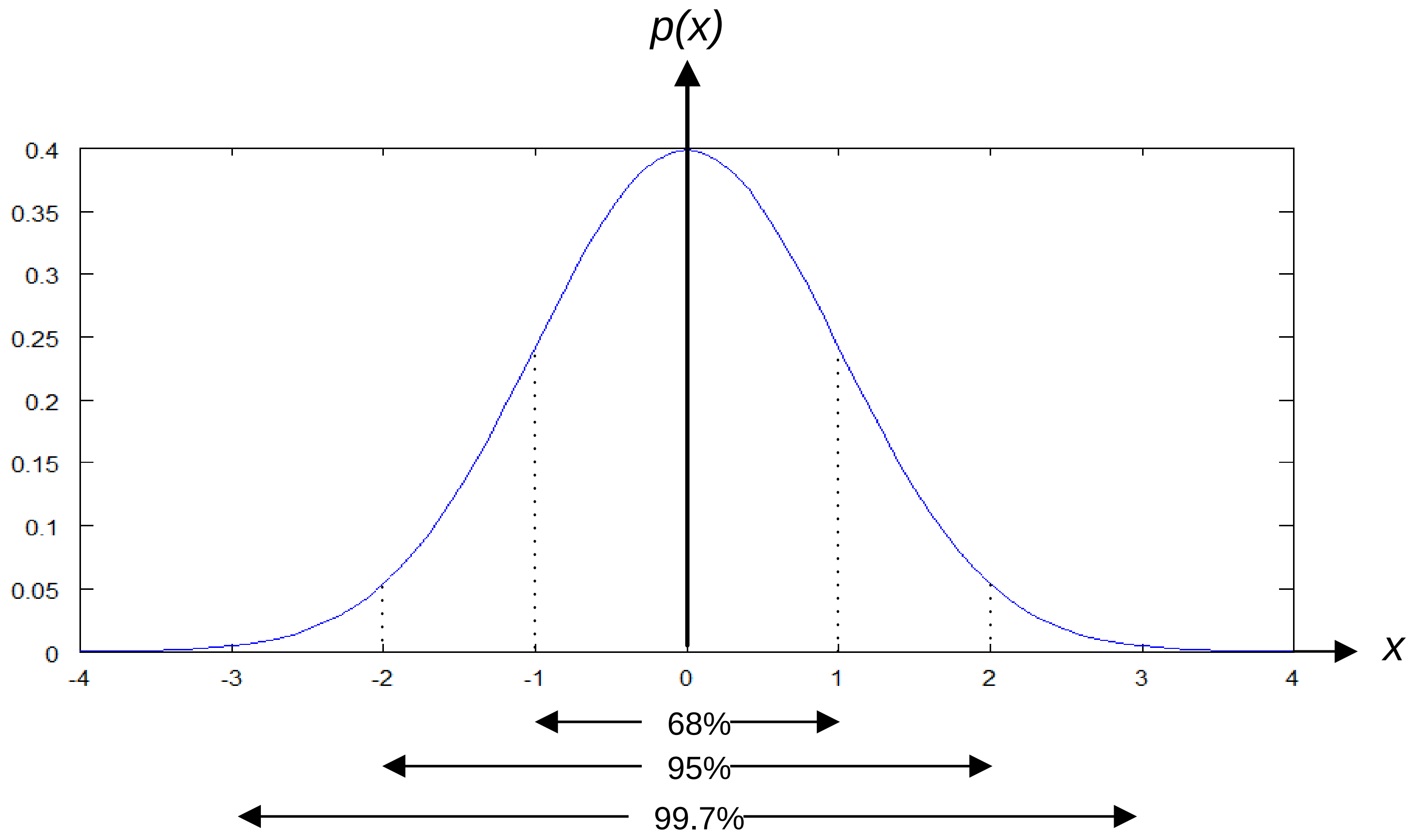


Figure 2: $p(x) \sim \mathcal{N}(0, 1)$ is shown. It has 68% of its probability mass in the range of $|x| \leq 1$ or, we can write $\Pr[|x-\mu| \leq \sigma] \approx 0.68$, 95% in the range $|x| \leq 2$ or, $\Pr[|x-\mu| \leq 2\sigma] \approx 0.95$ and 99.7% in the range $|x| \leq 3$ or, $\Pr[|x-\mu| \leq 3\sigma] \approx 0.997$.

Multivariate Normal Densities

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \left|\mathbf{\Sigma}\right|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

Quick Review

Bayes rule:

$$\Pr(Y \mid X) = \frac{\Pr(X \mid Y)\Pr(Y)}{\Pr(X)} \quad \Rightarrow \quad posterior = \frac{likelihood \times prior}{evidence}$$

Bayes rule for classification:

$$P(C_1 \mid x) = \frac{p(x \mid C_1)P(C_1)}{p(x)}$$

$$P(C_2 \mid x) = \frac{p(x \mid C_2)P(C_2)}{p(x)}$$

Naïve Bayes:

$$P(C_i \mid X) \equiv P(C_i)\prod_{j=1}^{d} p(x_j \mid C_i)$$

Multivariate Normal Densities

Assume that each of the *n* random variables x_i is normally distributed, each with its own *mean* and *variance*, as:

$$p_{x_i}(x_i) \sim N(\mu_i, \sigma_i^2)$$

Their joint density can be computed as:

$$p(x) = \prod_{i=1}^{n} p(x_i)$$

$$= \prod_{i=1}^{n} \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left\{ -\frac{(x_i - \mu_i)^2}{2\sigma_i^2} \right\}$$

$$= \frac{1}{(2\pi)^{n/2} \prod_{i=1}^{n} \sigma_i} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2} \right\}$$

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\mathbf{\Sigma}|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

Bernoulli Distribution

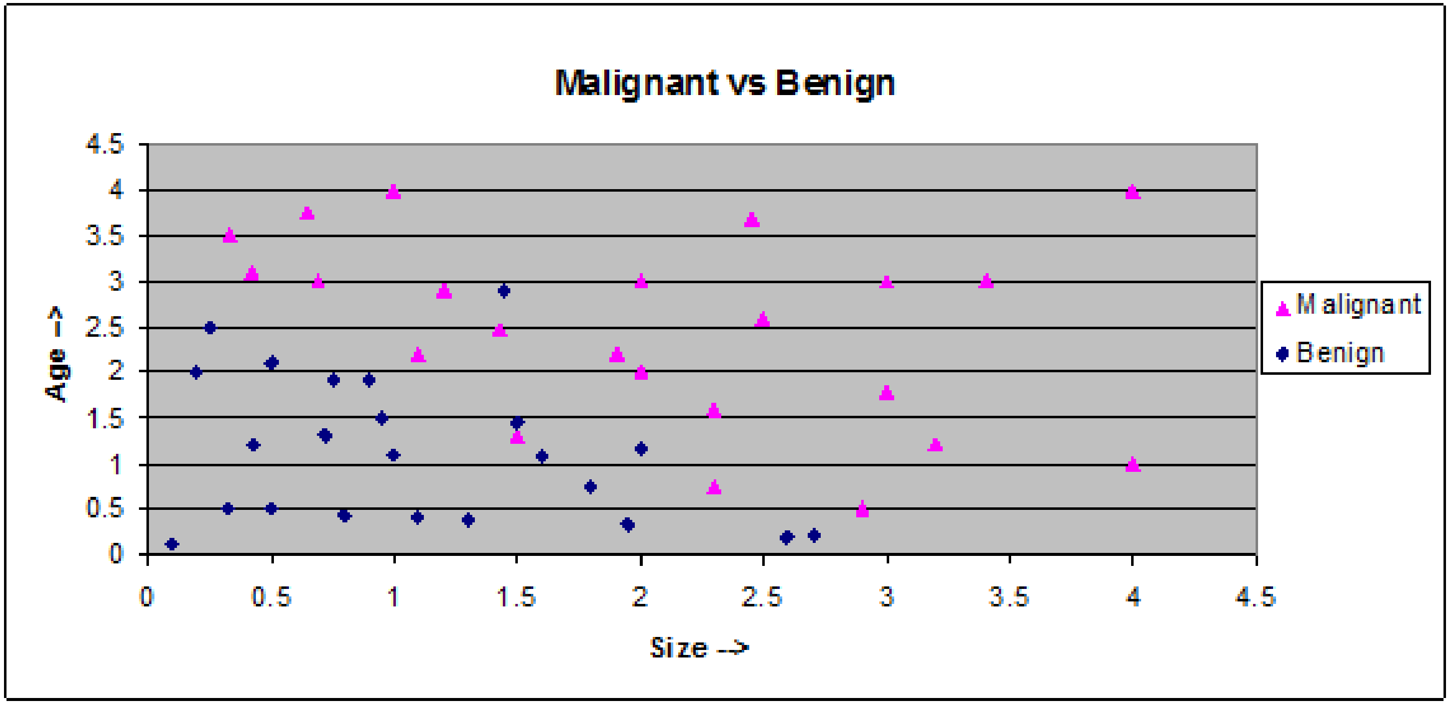
Bernoulli distribution is a discrete distribution which results two possible outcomes ($y \in \{0,1\}$) and one (say, y = 1) of them occurs with probability p whereas the other (y = 0, for example) occurs with probability (1-p). Such probability density function can be written as:

$$P(y) = \begin{cases} (1-p), \text{when } y = 0 \\ p, \qquad \text{when } y = 1 \end{cases}$$

In compact form, we can write it:

$$P(y) = p^{y}(1-p)^{(1-y)}$$

Gaussian Discriminant Analysis



$C_i \in \{0,1\}$, where '0' = Benign, '1'=Malignant.

$C_i \sim Bernoulli(p)$

$$=> \quad P(C_i) = p^{C_i}(1-p)^{(1-C_i)}$$

Gaussian Discriminant Analysis …

The pdfs for the classes can be expressed as;

$$p(X \mid C_i = 0) = \frac{1}{(2\pi)^{n/2} |\mathbf{\Sigma}|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_0)^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_0) \right\}$$

$$p(X \mid C_i = 1) = \frac{1}{(2\pi)^{n/2} |\mathbf{\Sigma}|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_1)^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_1) \right\}$$

GDA …

With $\boldsymbol{\mu}_0 = \begin{bmatrix} 1.1052 \\ 1.1286 \end{bmatrix}$ and $\boldsymbol{\Sigma} = \begin{bmatrix} 1.09229995 & 0.03073048 \\ 0.03073048 & 1.29917801 \end{bmatrix}$, the bivariate normal distribution for Benign class would look like

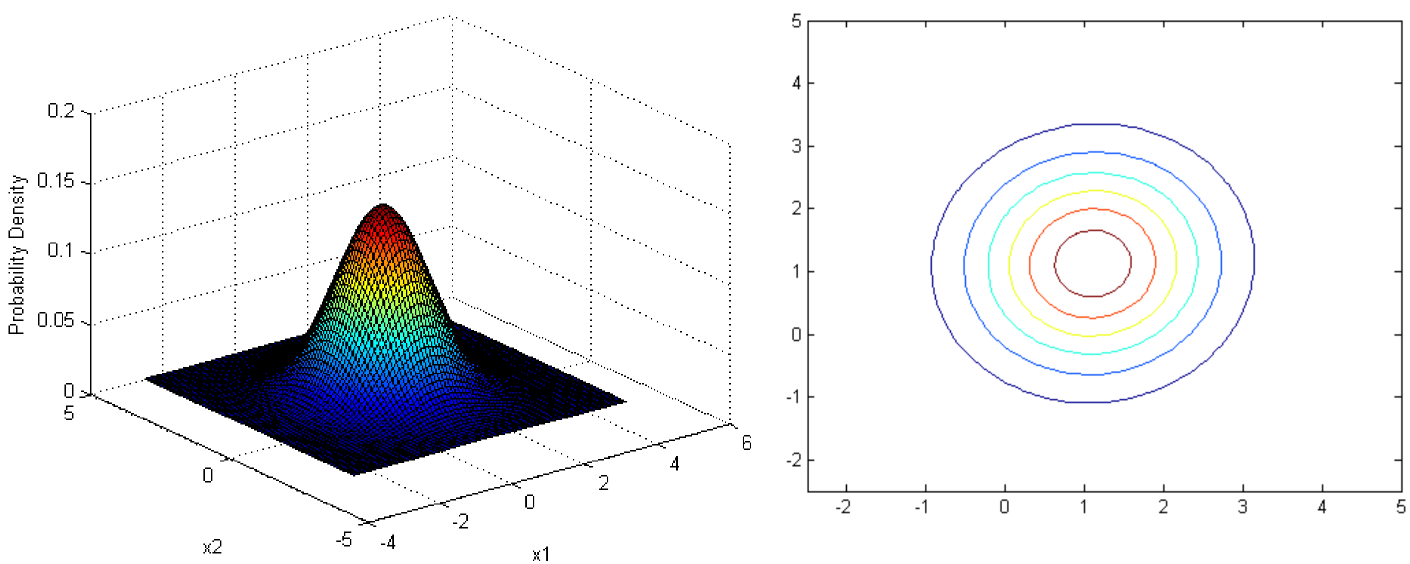
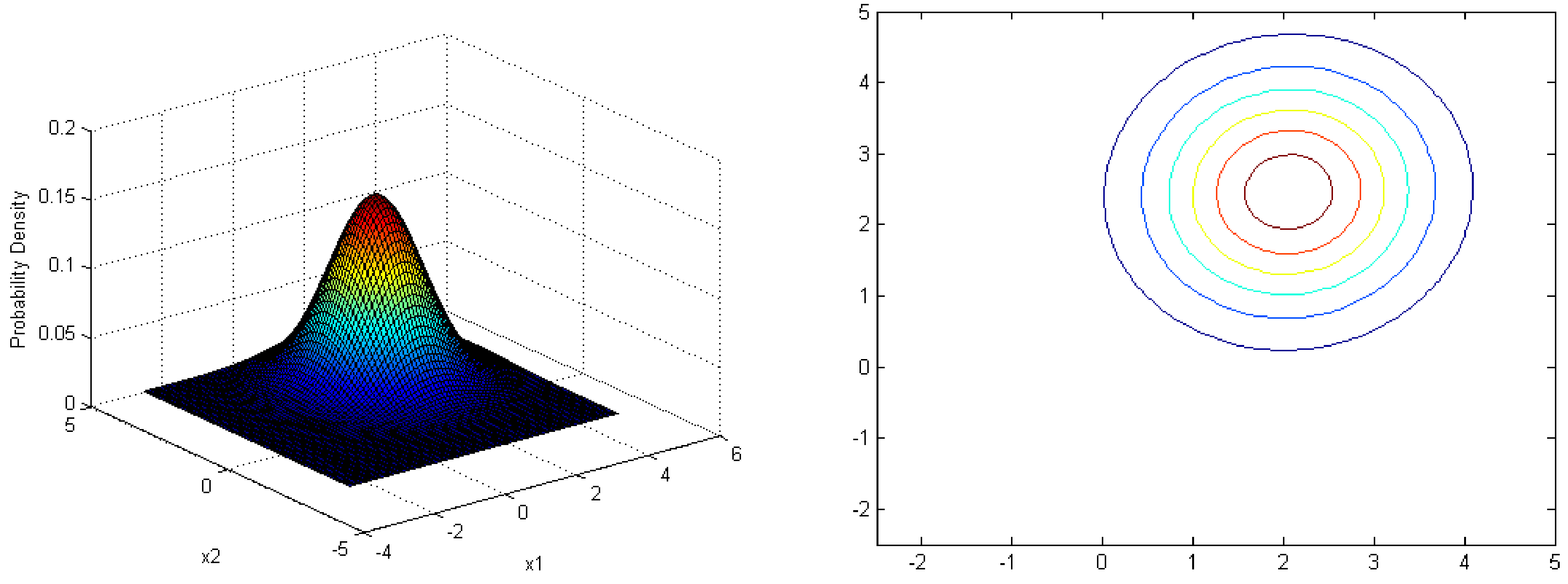


Figure 5: For benign class (a) pdf and (b) contour are shown.

GDA …

With $\boldsymbol{\mu}_1 = \begin{bmatrix} 2.0552 \\ 2.4578 \end{bmatrix}$ and $\boldsymbol{\Sigma} = \begin{bmatrix} 1.09229995 & 0.03073048 \\ 0.03073048 & 1.29917801 \end{bmatrix}$, the bivariate normal distribution for Malignant class would look like



(a) (b)

Figure 6: For malignant class (a) pdf and (b) contour are shown.

GDA ...

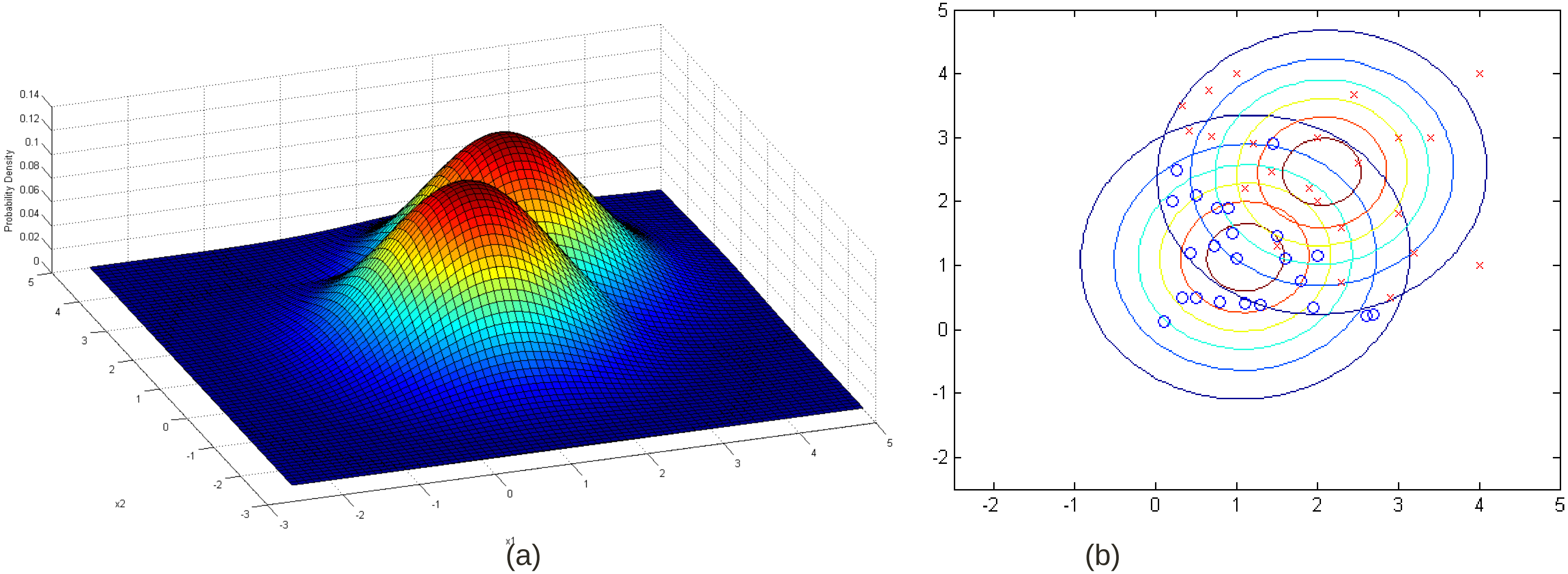


Figure 7: (a) Pdfs of both the classes: benign and malignant. (b) Given the datasets the Contours of both the classes are superimposed. '×' indicates data that belongs to malignant class and 'o' indicates data that belongs to benign class.