Data Mining and Machine Learning

Assignment Project Exam Help

https://eduassistpro.github.io/

Statistical

Add WeChat edu_assist_pro

Peter Jančovič

Objectives

- Review basic statistical modelling
- Review the notions of probability distribution and probability density function (PDF)
- Gaussian PD
- Multivariate Gaussian PD
- Parameter estimation for Gaussian PDFs

Discrete random variables

- Suppose that Y is a *random variable* which can take any value in a discrete set $X=\{x_1,x_2,\ldots,x_M\}$

- Suppose that y ,y ,…,y are samples of the random variable Y

- If c_m is the number of time $n = x_m$ then an

estimate of the probability n es the value x_m is given by:

$$P(x_m) = P \qquad \frac{\quad}{N}$$

Continuous Random Variables

- In most practical applications the data are not restricted to a finite set of values – they can take any value in *N*-dimensional space
- Counting the s of each value is no longer a viable way of probabilities…
- …but generalisations of th h are applicable to continuous variables – <u>non-parametric methods</u>

UNIVERSITY OF BIRMINGHAM

Continuous Random Variables

- An alternative is to use a <u>parametric</u> model
- Probabilities are defined by a small set of <u>parameters</u>

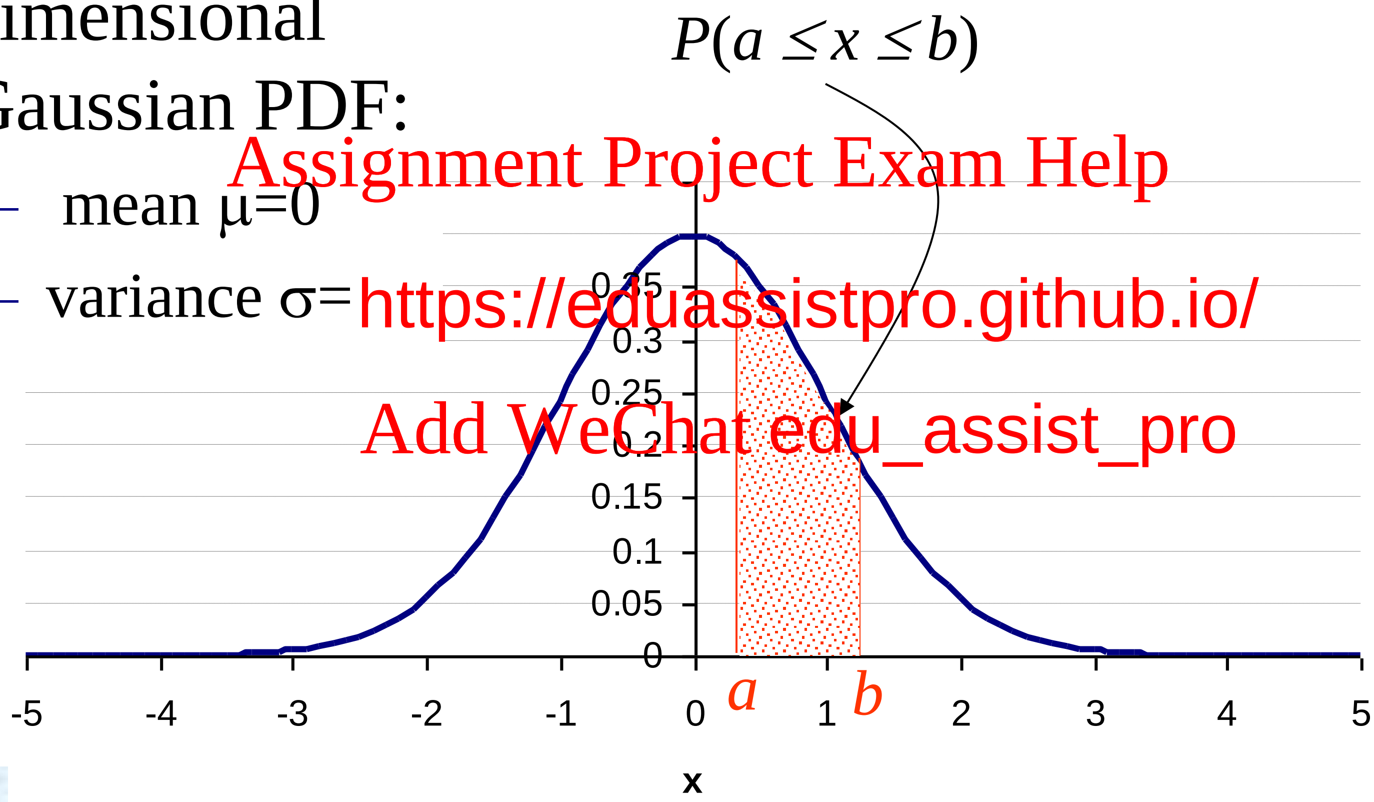
- Familiar exa … <u>aussian</u> model

- A (scalar/uni … ability density function (PDF) is defined … ameters – its <u>mean</u> μ and <u>variance</u> σ

- For a multivariate Gaussian PDF defined on a vector space, μ is the <u>mean vector</u> and σ is the <u>covariance matrix</u>

Gaussian PDF

- ‘Standard’ 1-dimensional Gaussian PDF:
  - mean μ=0
  - variance σ=



Gaussian PDF

- For a 1-dimensional Gaussian PDF *p* with mean μ and variance σ:

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

Constant to ensure area under curve is 1

Defines 'bell' shape

Standard Deviation

- Standard deviation is the square root of the variance
- For a Gaussian PDF:

  – 68% of th lies within one standard

  ean

  – 95% of the area under t es within two s.ds of the mean

  – 99% of the area under the curve lies within three standard deviations of the mean

Standard Deviation

- In other words, if $s = \sqrt{\sigma}$ then:

$$P(\mu - s \leq x \leq \mu + s) = 0.68$$

$$P(\mu - 2s \leq x \leq \mu + 2s) = 0.95$$

$$P(\mu - 3s \leq x \leq \mu + 3s) = 0.99$$

Multivariate Gaussian PDFs

- A (univariate) Gaussian PDF assumes the random variable takes scalar values
- In the case where the random variable takes N dimensional ... esponding PDF is called a mult ... and is given by:

$$p(x)=\frac{1}{\sqrt{(2\pi)^N|\Sigma|}}\exp\left(\frac{\quad}{2}\left(\quad {}^{1}(m-x)\right)\right)$$

UNIVERSITY OF BIRMINGHAM

Visualising multivariate Gaussian PDFs

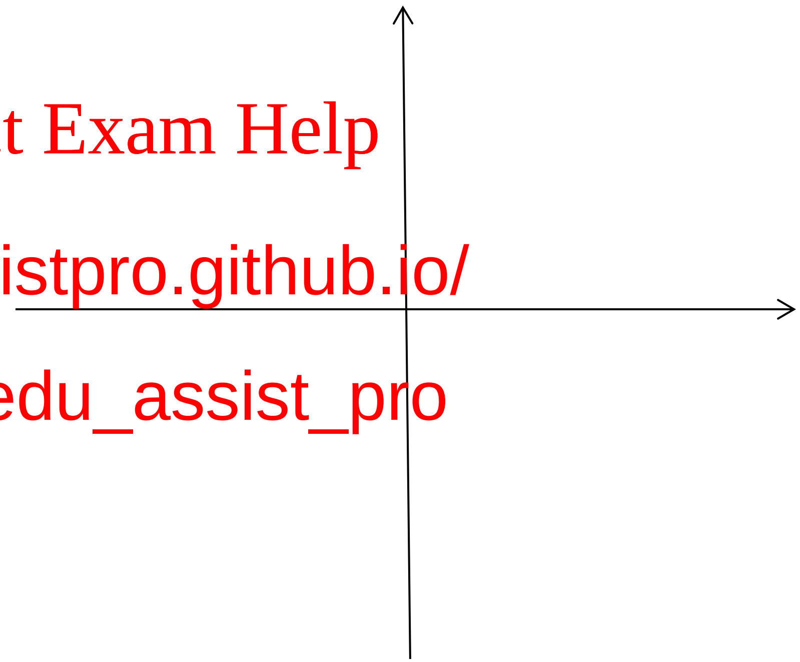
- It's easy to sketch a 1 dimensional Gaussian PDF, using the rules about the proportion of the area that lies within 1, 2 and 3 standard deviations of the mean and the value for $p(\mu)$.

- 2D Gaussian PDFs can be plotted using MATLAB's 3D plotting functions
- A simpler way to visualise a 2D Gaussian PDF is to plot the 1 standard deviation contour. This is the set of points that lie 1 standard deviation from the mean

Example

- If $\Sigma = \begin{bmatrix} 9 & 0 \\ 0 & 4 \end{bmatrix}$, standard deviations in 'x' and 'y' directions ar

  respectively, and the 1 s.d. contour is an ellipse:



Example 2:

- Now suppose $\Sigma = \begin{bmatrix} 7.75 & 2.17 \\ 2.17 & 5.25 \end{bmatrix}$ and $m = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$

- Calculate th ition of Σ

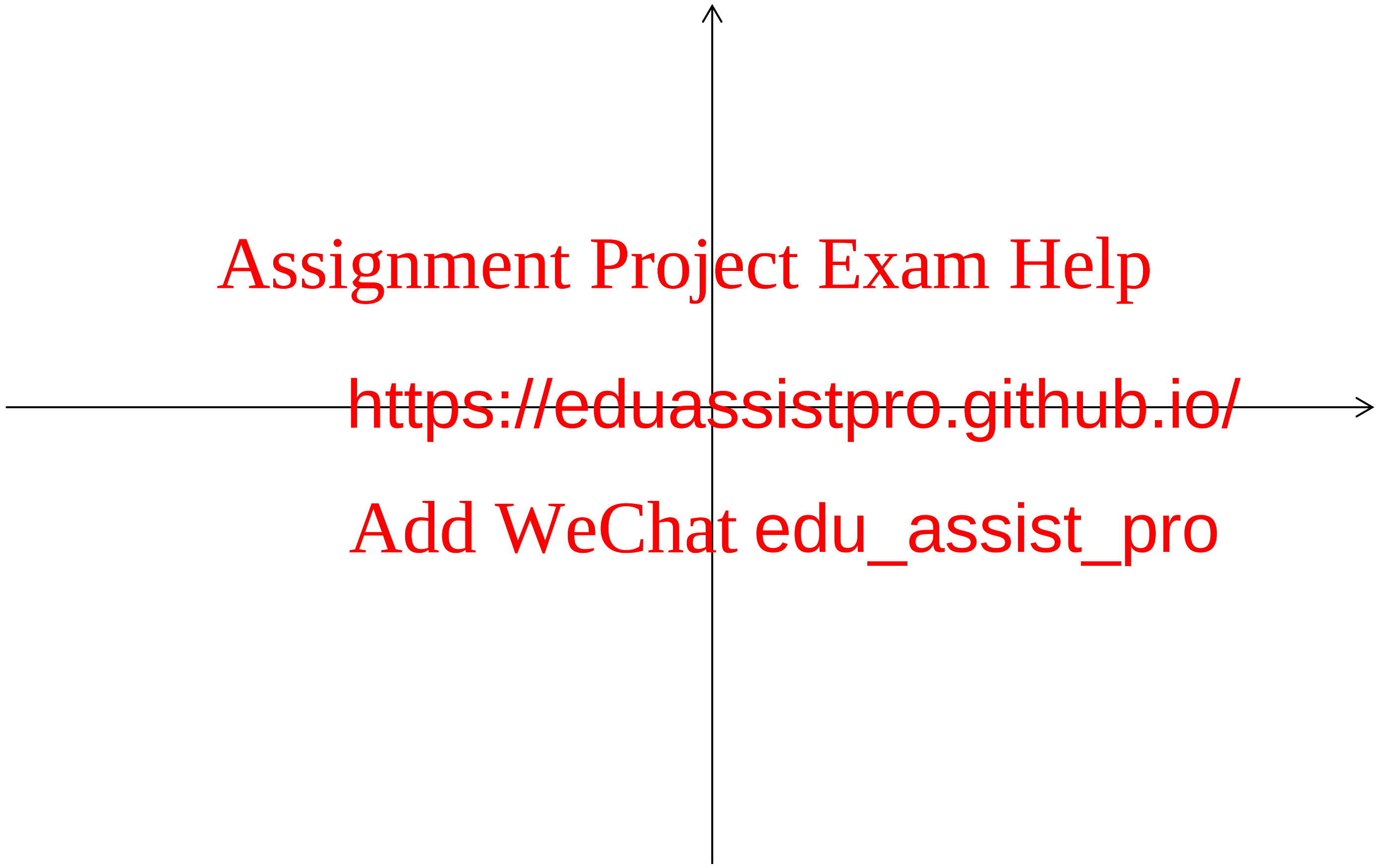
$$\Sigma = UDU^T = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 9 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ \frac{-1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}$$

Example 2 (continued)

- Note U is a rotation through 30^o
- Hence the one standard deviation contour is the same as in the previous example, but rotated through 30^o and trans

$$m = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

Example 2 (continued)



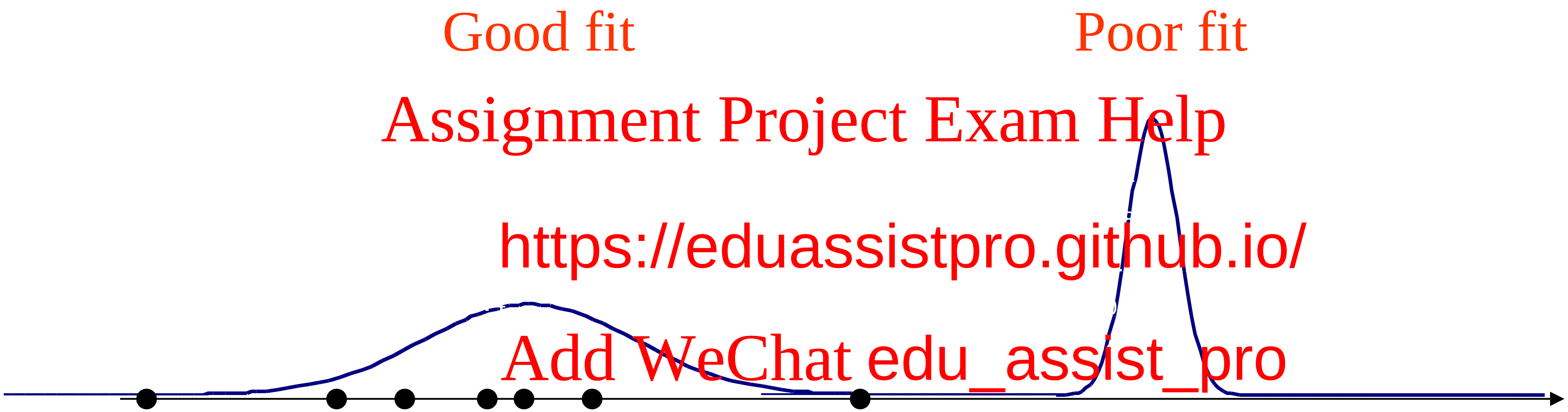
Fitting a Gaussian PDF to Data

- Suppose $y = y_1, \ldots, y_t, \ldots, y_T$ is a set of T data values
- For a Gaussian PDF p with mean μ and variance σ, define:

$$p(y \mid \mu, \sigma) = \prod_{t=1}^{T} p(y_t \mid \mu, \sigma)$$

- How do we choose μ and σ to maximise $p(y|\mu, \sigma)$?

Fitting a Gaussian PDF to Data



Maximum Likelihood Estimation

- The ‘best fitting’ Gaussian maximises $p(y|\mu,\sigma)$
- Terminology:
  - $p(y|\mu,\sigma)$, as a function of y is the <u>probability</u> (<u>density</u>)
  - $p(y|\mu,\sigma)$, <u>kelihood</u> of μ,σ
- Maximising $p(y|\mu,\sigma)$ with respect to μ,σ is called <u>Maximum Likelihood (ML)</u> estimation of μ,σ

ML estimation of μ,σ

- Intuitively:
  - The maximum likelihood estimate of μ should be the average value of $y_1,\ldots,y_T$. (the sample mean)
  - The maxi te of σ should be the variance of $y_1,\ldots,y_T$. le variance)

- This is true: $p(y|\ \mu,\ \sigma)$ is by setting:

$$\mu = \frac{1}{T}\sum_{t=1}^{T} y_t, \quad \sigma = \frac{1}{T}\sum_{t=1}^{T}\left(y_t\right.$$

Proof

First note that maximising $p(y)$ is the same as maximising $\log(p(y))$

$$\log p(y \mid \mu, \sigma) = \log \prod_{t=1}^{T} p(y_t \mid \mu, \sigma) = \sum_{t=1}^{T} \log p(y_t \mid \mu, \sigma)$$

Also

$$\log p(y_t \mid \mu, \sigma) = -\frac{1}{2}\log(2\pi\sigma) - \frac{(\mu - y_t)^2}{\sigma}$$

At a maximum:

$$0 = \frac{\partial}{\partial \mu} \log p(y \mid \mu, \sigma) = \sum_{t=1}^{T} \frac{\partial}{\partial \mu} \log p(y_t \mid \mu, \sigma) = \sum_{t=1}^{T} \frac{-2(\mu - y_t)(-1)}{\sigma}$$

So, $T\mu = \sum_{t=1}^{T} y_t, \mu = \frac{1}{T}\sum_{t=1}^{T} y_t$

Multi-modal distributions

- In practice the distributions of many naturally occurring phenomena do not follow the simple bell-shaped Gaussian curve

- For example several difference so veral distinct peaks (e.g. distribution of adults)

- These peaks are the <u>modes</u> of the distribution and the distribution is called <u>multi-modal</u>

Summary

- Reviewed basic statistical modelling, probability distribution, probability density function
- Gaussian PDFs
- Multivariate G
- Maximum like timation

- In the next session we will introduce Gaussian mixture PDFs (GMMs) and ML parameter estimation for GMMs