High dimensional data

- Volume
- Ratios
- Axes

Normal Distribution

\[ f(x) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]

\( M: \text{mean} \)
\( \sigma^2: \text{variance} \)
\( \mu = \text{mean} \)
\( \sigma^2 = \text{variance} \)

Standard normal
\( \mu = 0, \sigma = 1 \)

Confidence intervals
\( P \left( \hat{\mu} - \delta \leq \mu \leq \hat{\mu} + \delta \right) \geq 0.95 \)

\( \bar{X} \) is distributed normally
\( E[\bar{X}] = \mu \)

Gaussian
Normal distribution: parametric distribution \& choice by default

\( D = \begin{bmatrix} x_1 & x_2 & \cdots & x_d \end{bmatrix} \) is \( d \)-dim data
d-dim data
PDF or prob density function for \( \mathbf{X} \)
\[
\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{pmatrix}
\]

\[f(x) = N_d(x)\]
A normal distribution to estimate the PDF

mixture of 2 normals

\[
f(x) = \omega_1 N_1(x) + \omega_2 N_2(x)
\]

Even do a mixture of \( n \) Gaussians, one per point

\[
f(x) = \omega_1 N_1(x | \mu_1, \Sigma_1) + \omega_2 N_2(x | \mu_2, \Sigma_2)
\]
1D normal $N(\mu, \sigma^2)$

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

$d$ Dim Normal $N(\mu, \Sigma)$

$$f(x) = \frac{1}{(2\pi)^{d/2} \sqrt{|\Sigma|}} e^{-\frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu)}$$

If $\Sigma = \Sigma^{-1}$ sample

$$-\frac{(x-\mu)^T \Sigma (x-\mu)}{2}$$

Mahalanobis distance

distance between $\tilde{x}$ and $\tilde{\mu}$

$$\Delta^2(\tilde{x}, \tilde{\mu}) = (\tilde{x} - \tilde{\mu})^T \Sigma^{-1} (\tilde{x} - \tilde{\mu})$$

Euclidean distance

$$\|\tilde{x} - \tilde{\mu}\|_2^2 = (\tilde{x} - \tilde{\mu})^T (\tilde{x} - \tilde{\mu})$$

Unweighted

$$= (\tilde{x} - \tilde{\mu})^T \Sigma^{-1} (\tilde{x} - \tilde{\mu})$$

$\Sigma = \Sigma^{-1}$

estimate from the D

Unknown

High dimensional normals

$d: \text{dim mean}$

$d: \Sigma: \text{covariance matrix}$

$$d \times d \rightarrow O(d^2)$$
\[ d \times d \rightarrow o(d^2) \]

Can become an issue

If \( n \) is small, hard to get a good estimate of \( \Sigma \)

\( \Sigma \) is not full rank

fewer than \( d \) non-zero eigenvalues

\( \Sigma^{-1} \) does not exist

Use pseudo-inverse \( \Sigma^+ \)

Inverse of the truncated eigen decomposition of \( \Sigma \)

(over non-zero \( \lambda_i \))

\[
\Sigma = \sum_{i=1}^{d} \lambda_i u_i u_i^T
\]

Standard multivariate Normal

\[
\mu = 0
\]

\[
\Sigma = I \leftarrow \text{hyper-sphere}
\]

\[
f(\mathbf{x}) = \frac{1}{(N^{2\pi})^{\frac{d}{2}}} e^{-\frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x}}
\]
\( \mathbf{x} \in \mathbb{R}^d \)

Peak density \( \mathbf{z} = \mathbf{0} \)

\[
f(\mathbf{0}) = \frac{1}{(N^{2\pi})^{\frac{d}{2}}}
\]
Q: Find all points within a fraction \( \alpha \) of the peak density.

Density or \( z \) = \( f(\tilde{z}) \)

Peak density = \( f(\tilde{0}) \)

\[
\frac{f(\tilde{z})}{f(\tilde{0})} \geq \alpha
\]

\( \alpha = 0.5 \)

LD: 76.1 \% of the points lie in this region.

d=2: 50 \%.

d=3: 29.7 \%.

d \to \infty ?

\( \alpha \to 0.5 \)
\[ \text{find } x \quad \frac{f(x)}{f(\bar{x})} \geq \alpha \]

\[ \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\bar{x})^2} \geq \alpha \]

\[ \frac{e^{-\frac{1}{2}(x-\bar{x})^2}}{\alpha} \geq 1 \]

\[ -\frac{(x-\bar{x})^2}{2} \geq \ln \alpha \]

\[ (x-\bar{x})^2 \leq -2\ln \alpha \]

What is the probability of finding a point within a fraction of the peak?

\[ \lim_{d \to \infty} P \left( \frac{f(x)}{f(\bar{x})} \geq \alpha \right) = \lim_{d \to \infty} P \left( \frac{(x-\bar{x})^2}{2\sum_{i=1}^{d} x_i^2} \leq -2\ln \alpha \right) \]

\[ \frac{\text{sum of } d \text{ square}}{\sum_{i=1}^{d} x_i^2} \]

\[ \frac{x-\bar{x}}{\sqrt{d}} \text{ follows a } \chi^2 \text{ distribution with } d \text{ degrees of freedom} \]

\[ \lim_{d \to \infty} P \left( \chi^2_d \leq -2\ln \alpha \right) \to \]

\[ \alpha > 0 \]

\[ \alpha = \text{50%} \]

\[ \Rightarrow \text{as } d \to \infty, \text{ there are no points in the 50% region} \]
$\alpha = 0.1$

$\alpha = 0.01$

$\rho (\sqrt{\frac{d}{D^2}} \leq 2 \sin \alpha) \Rightarrow 0$

Kernel methods

Exam I:

full: chapter 1, 2, 6

partial: chapter 7 (7.1 - 7.2)

Live Exam 10:00 am
12:00 pm

10 min to submit on submitly.
Exam will be posted before 12:10 pm.

Career fair: Email me today! (email temp server)
8:00 am - 10:00
Current time: Email me today! (email server)
8:30 am → 10:30
10 min to submit

Accommodations: Email me today.

Practice: Q on each chapter
discuss sol on campuswire

Open book

Assign 1: Any issue contact TA by tomorrow!
finalized after that.

Kernel Methods

D

Point view

Attribute view

Graph/similarity view

pairwise similarity
(no vector space)

K(x, y)

"good" similarity

Kernel function

Vectors in $\mathbb{R}^d$ or $\mathbb{R}^n$

Similarity

K(x, y)

K(x, y)

K(x, y)
Kernel function

\[ \mathbf{x}_i \to \mathbf{y}_j \to \text{some abstract vector space, such that} \]
\[ k(\mathbf{x}_i, \mathbf{x}_j) = \text{dot product in kar space.} \]

Complex Data

Image

Video (seq of frames)

Text

Multimodal (text, image, hyperlinks, etc.)

Set of images

How do you construct feature vector \( \mathbf{z} \)?

Image \( 1000 \times 1000 \)

\( \rightarrow \) \( 10^6 \) dim vector of

Pixel values

Image \( i \)

Pixel level vector

Semantic vector

Feature engineering \( \rightarrow \) Deep learning (Convolution)

\[ k(\text{Image } i, \text{Image } j) \rightarrow \text{scalar value} \]

Input space

Input objects
e.g. Image

\( \rightarrow \)

Underlying feature space
e.g. Image

\[ k(x_i, x_j) \]

Scalar

Conceptually

\[ \phi(x_i)^T \phi(x_j) \]

a dot product in \( F \)

Never explicitly constructed

Linear in \( \phi^2 \)

High-dimensional space