Research Methods

Lecturer: Steve Maybank

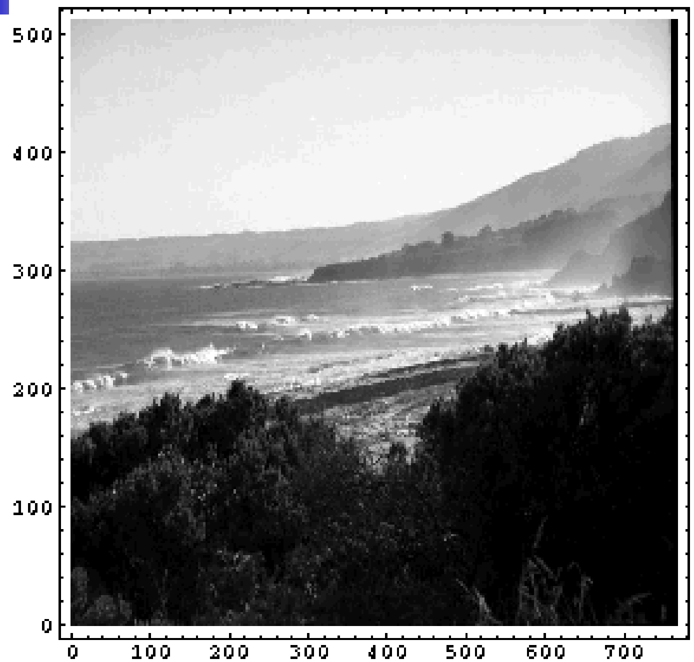
Department of Computer Science and Information Systems

sjmaybank@dcs.bbk.ac.uk

Autumn 2017

Data Research Methods in Computer Vision

Digital Images



Original colour image from the Efficient Content Based Retrieval Group, University of Washington

95	110	40	34
125	108	25	91
158	116	59	112
166	132	101	124

A digital image is a rectangular array of pixels. Each pixel has a position and a value.

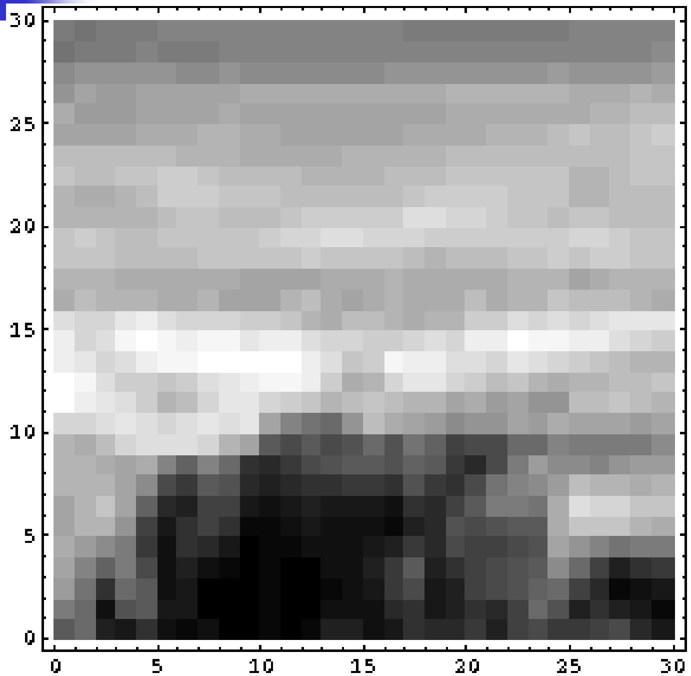
Size of Images

- Digital camera, 5,000x5,000 pixels, 3 bytes/pixel -> 75 MB.
- Surveillance camera at 25 f/s ->
  1875 MB/s.
- 1000 surveillance cameras -> ~1.9 TB/s.
- Not all of these images are useful!

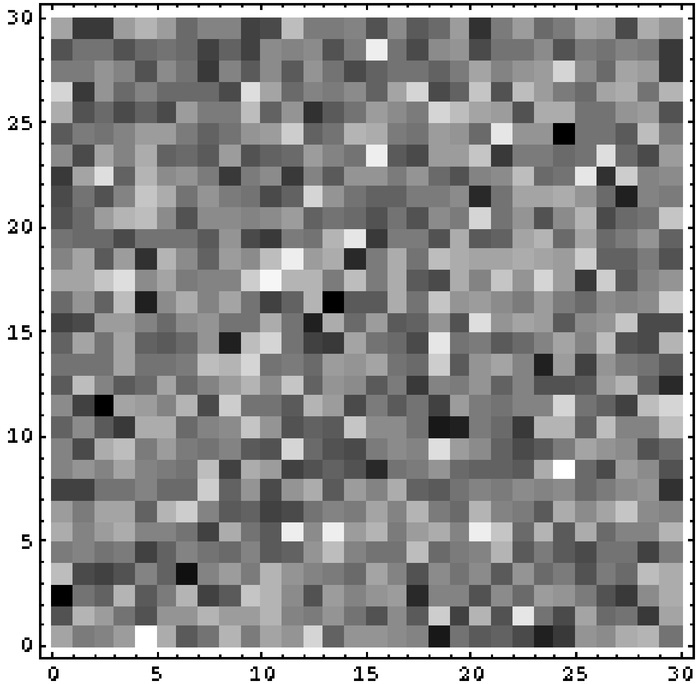
Image Compression

- Divide the image into blocks, and compress each block separately, e.g. JPEG uses 8x8 blocks.
- Lossfree compression: the original image can be recovered exactly from the compressed image.
- Lossy compression: the original image cannot be recovered.

Why is Compression Possible?

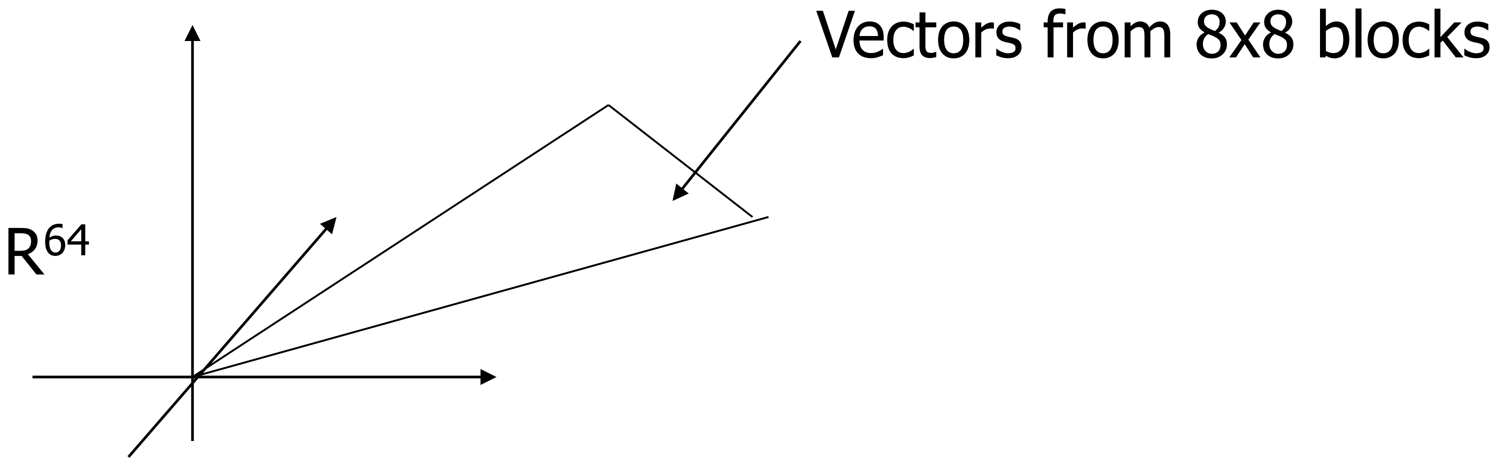


Natural image: values of neighbouring pixels are strongly correlated.



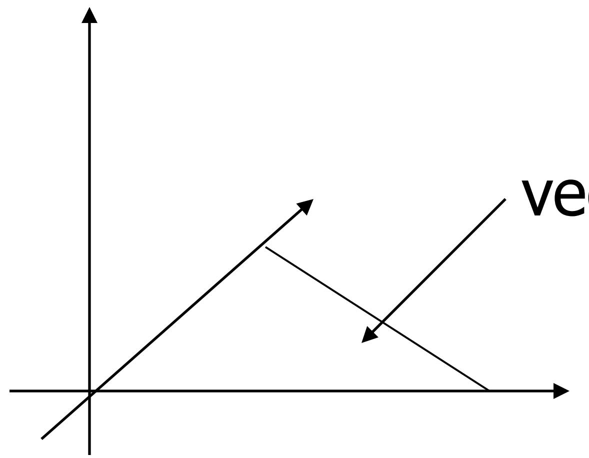
White noise image: values of neighbouring pixels are not correlated. Compression discards information.

Measurement Space



Each 8x8 block yields a vector in R^{64}. The vectors from natural images tend to lie in a low dimensional subspace of R^{64}.

Strategy for Compression



Choose a basis for R^{64} in which the low dimensional subspace is spanned by the first few coordinate vectors. Retain these coordinates and discard the rest.

Discrete Cosine Transform

Let $w \in R^{64}$ be a vector obtained from an 8×8 block. Then

$$\text{DCT}(w) = Uw$$

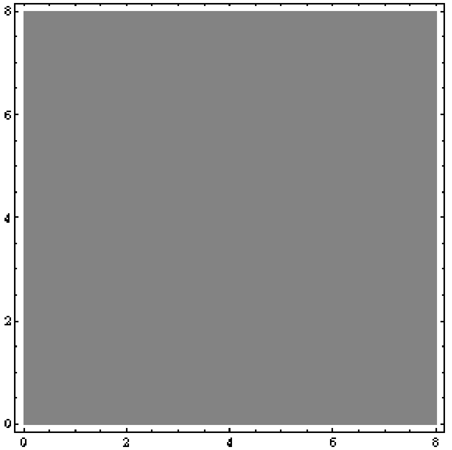
where U is a certain 64×64 othogonal matrix, $U^T U = I$. Note that $\|\text{DCT}(w)\| = \|Uw\| = \|w^T U^T U w\|^{1/2} = \|w\|$, where $\|.\|$ is the Euclidean norm.

Define vectors $e(i) \in R^{64}$, by $e(i)_j = 0,\ j \neq i, e(i)_i = 1$. Then

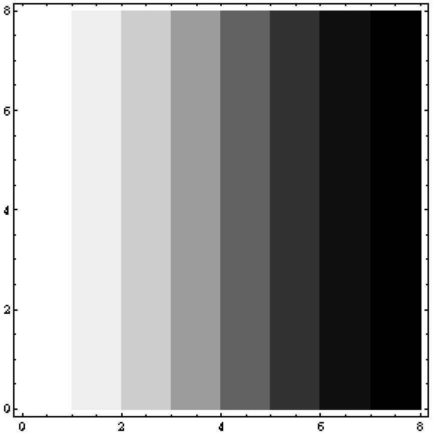
$$\text{DCT}(w) = \sum_{i=1}^{64} c_i e(i) \quad \text{and} \quad w = \sum_{i=1}^{64} c_i U^T e(i).$$

If i is large, then $|c_i|$ tends to be small.

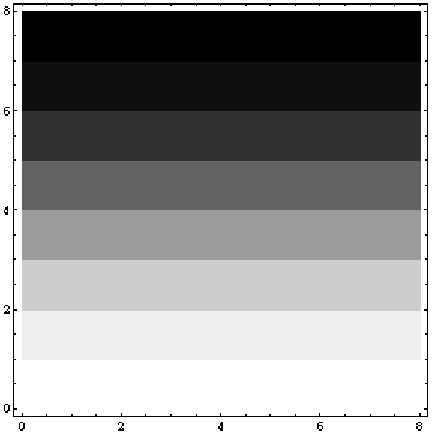
Basis Images for the DCT



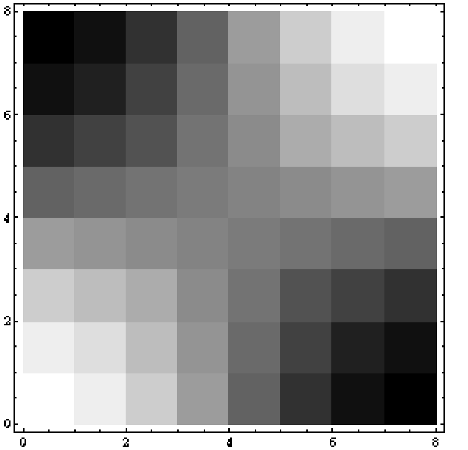
$U^T e(1)$



$U^T e(2)$

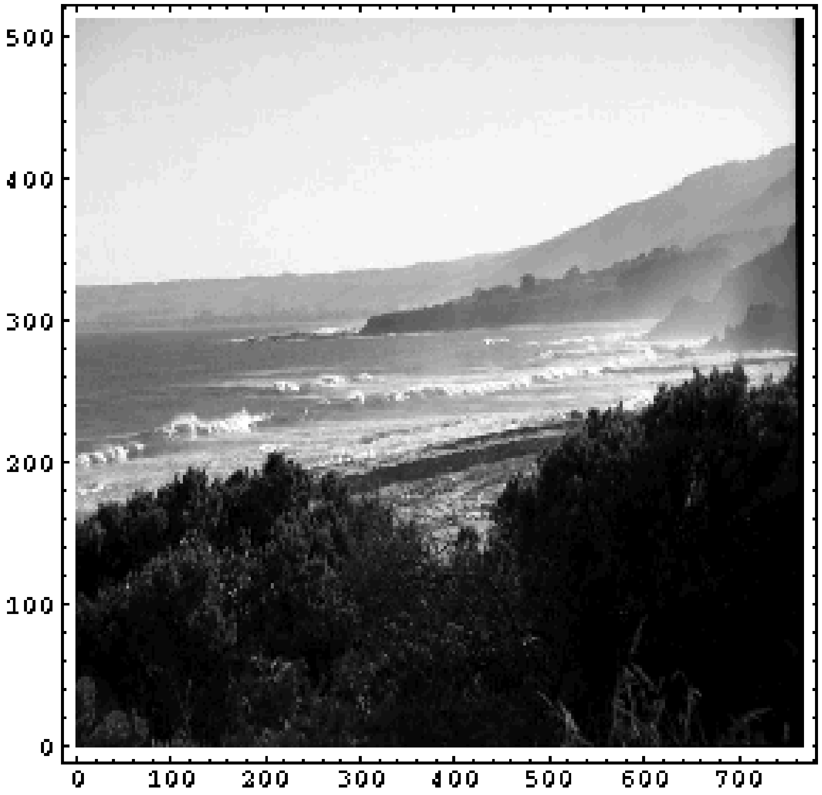


$U^T e(3)$



$U^T e(4)$

Example of Compression using DCT



Original image

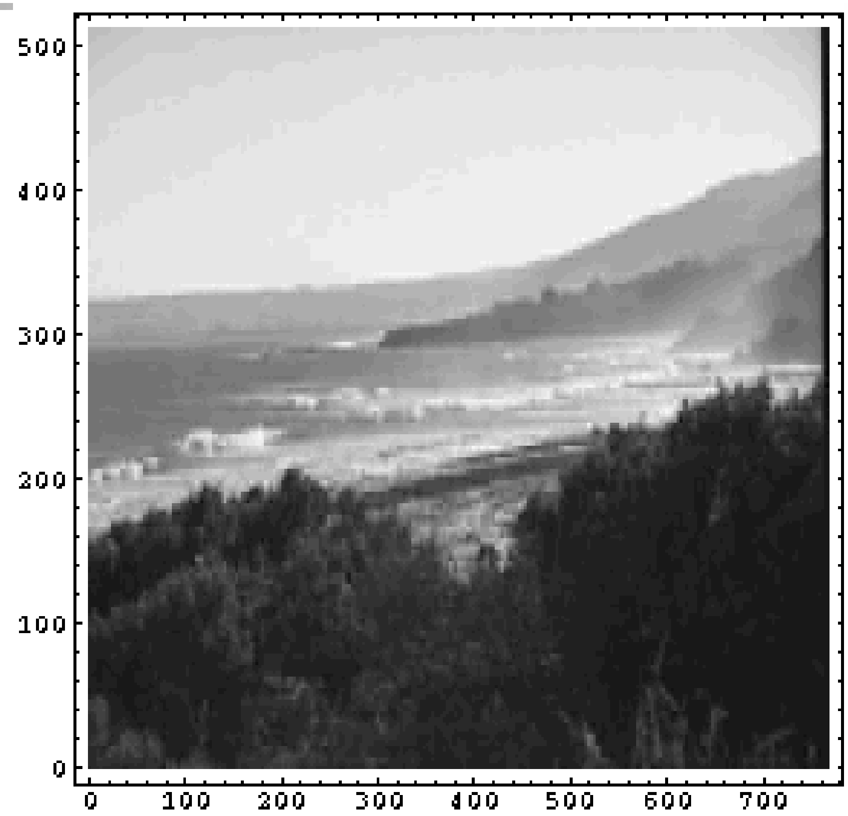
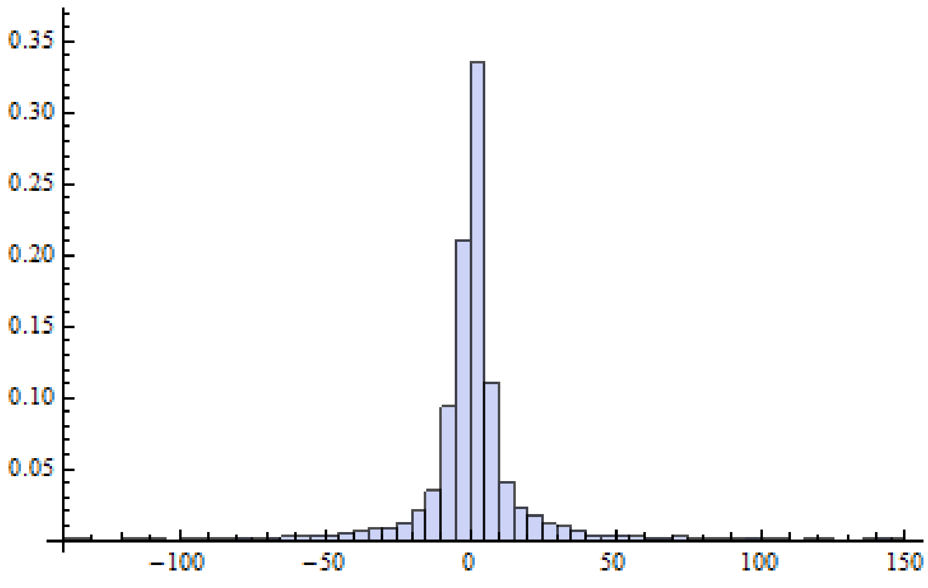


Image constructed from 3 DCT coefficients in each 8x8 block.

Histogram of a DCT Coefficient



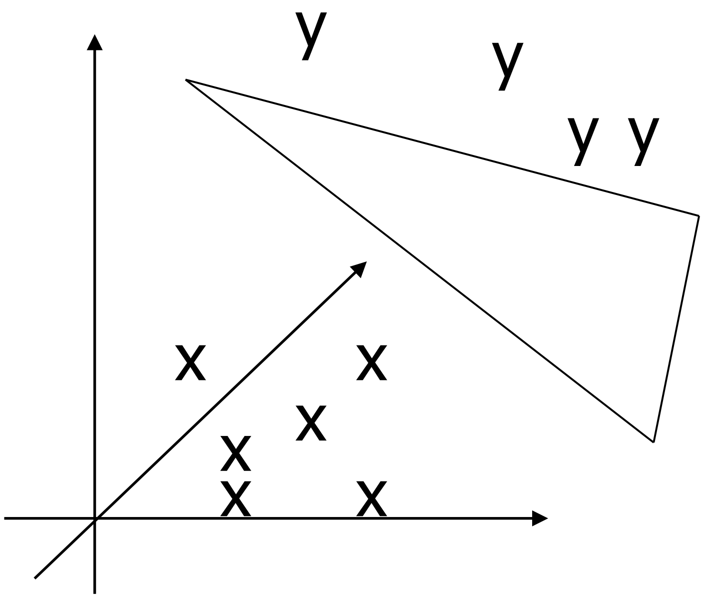
The pdf for ci is leptokurtic, i.e. it has a peak at 0 and "fat tails"

$$\mathrm{DCT}(w) = \sum_{i=1}^{64} c_i e(i)$$

Sparseness of the DCT Coefficients

- For a given 8×8 block, only a few DCT coefficients ci are significantly different from 0.
- For a given DCT coefficient, there exist some blocks for which it is large.

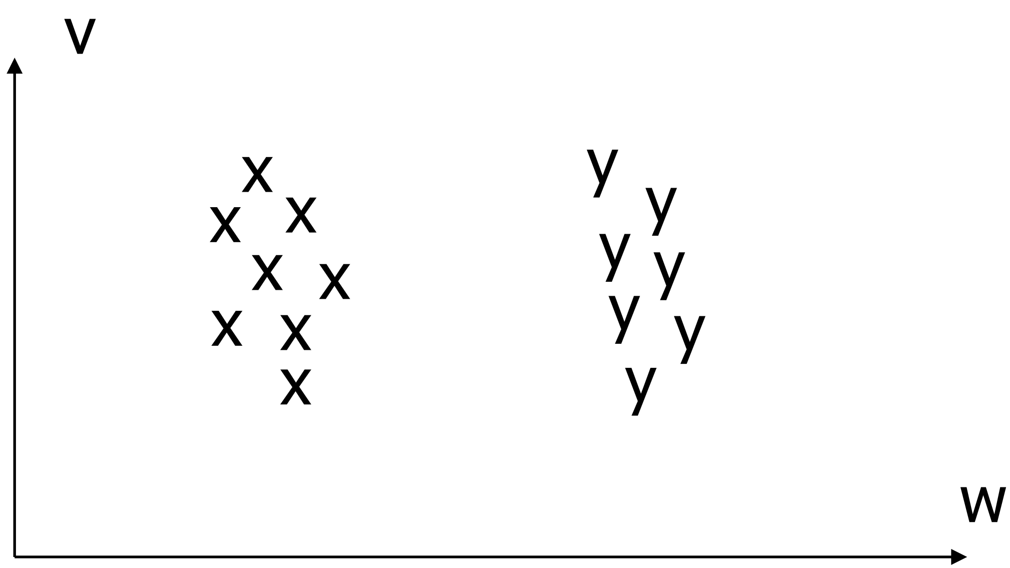
Linear Classification



Given two sets X, Y of measurement vectors from different classes, find a hyperplane that separates X and Y.

A new vector is assigned to the class of X or to the class of Y, depending on its position relative to the hyperplane.

Projection to a Line



Projection to the line defined by the unit vector w separates the two sets, $x \longmapsto x.w$

Fisher Linear Discriminant

Let $X_i, 1 \le i \le m$ and $Y_i, 1 \le i \le n$ be two sets of points in $\mathbb{R}^k$ from different classes.

Mean values: μ_X, μ_Y
Covariances: C_X, C_Y

Project the X_i and the Y_i onto the line with direction w, $X_i \mapsto w.X_i$, etc.

$$\frac{\text{between class variance}}{\text{within class variance}} = \frac{\left(w.(\mu_X - \mu_Y)\right)^2}{w^T(C_X + C_Y)w}$$

Maximise Ratio of Variances

Equate the derivative of the ratio with 0, to obtain

$$(C_X + C_Y)w = \lambda(\mu_X - \mu_Y)$$

where λ is an arbitrary number

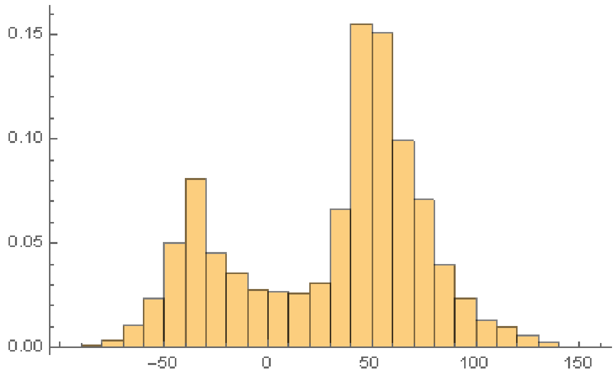
Two Classes of Edges

3x3 blocks matching mask
{{-1, 0, 1}, {-2, 0, 2}, {-1, 0, 1}}

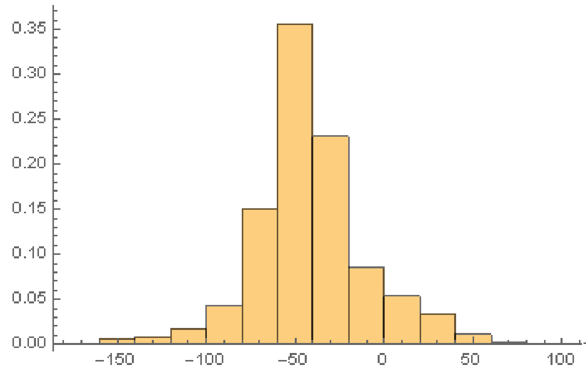
3x3 blocks matching mask
{{-1,-2,-1}, {0, 0, 0}, {1, 2, 1}}

$$x.mask > 0.8$$

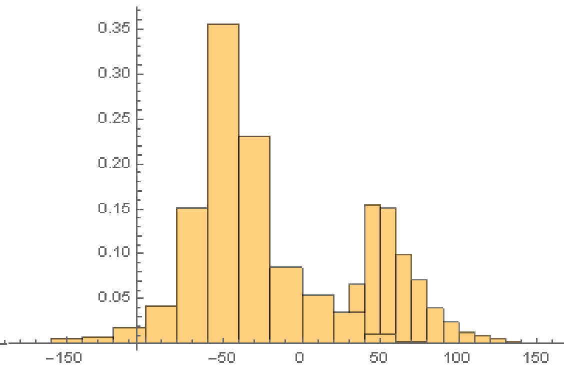
Projections Onto a 1-Dimensional FLD



Histogram for
{{-1,0,1}, {-2,0,2}, {-1,0,1}}



Histogram for
{{-1,-2,-1}, {0,0,0}, {1,2,1}}



Combined
histograms

Discrete Distribution

- A probability distribution on a discrete set S={1, 2,…, n} is a set of numbers p_i such that

$$0 \le p_i \le 1$$

$$\sum_{i=1}^{n} p_i = 1$$

Interpretations

- Bayes: p_i is a measure of our knowledge that item i is chosen from S.
- Frequentist: in a large number m of independent samples from S, i occurs approximately $m\,p_i$ times

Terminology

- Event: subset of S
- Probability of event E:

$$\mathrm{P(E)} = \sum_{i \epsilon E} p_i$$

- Conditional Probability:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Example

- Roll two dice. F=event that total is 8.
- S={(i, j), 1<=i, j<=6}
- The pairs (i, j) all have the same probability, thus

  P({i, j})=1/36, 1<=i<=36

- F={(6,2), (2,6), (3,5), (5,3), (4,4)}
- P(F) = 5/36

Example of a Conditional Probability

- E={(6,2)}. What is the probability of E given F (the total is 8)?

- $P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{P(E)}{P(F)} = \frac{\frac{1}{36}}{5/36} = 1/5$

Independent Events

- The events E, F are independent if

$$P(E \cap F) = P(E)P(F)$$

- Example: E=first number is 6

  F=second number is 5

$$P(E \cap F) = \frac{1}{36}$$

- P(E) = 1/6, P(F)=1/6

Bayes Theorem

- If E, F are two events then

$$P(E|F) = P(F|E)P(E)/P(F)$$

- Example: roll two dice

  E = sum is 7

  F = {(4, 3)}

  P(E|F) = 1, P(E) = 1/6,

  P(F) = 1/36, P(F|E) = 1/6

Probability Density Function

- A pdf of the real line R is a function

$$f: R \to R$$

such that

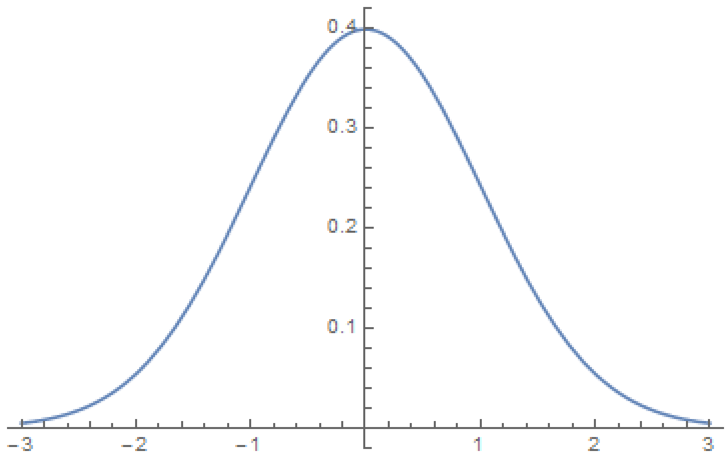
$$f(x) \geq 0, x \epsilon R$$

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

- A pdf is used to assign probabilities to subsets of R:

$$P(A) = \int_A f dx$$

The Gaussian PDF



$$f(x) = (2\pi)^{-1/2} e^{-x^2/2}$$

Mean value: $\mu = \int_{-\infty}^{\infty} x f(x) dx$

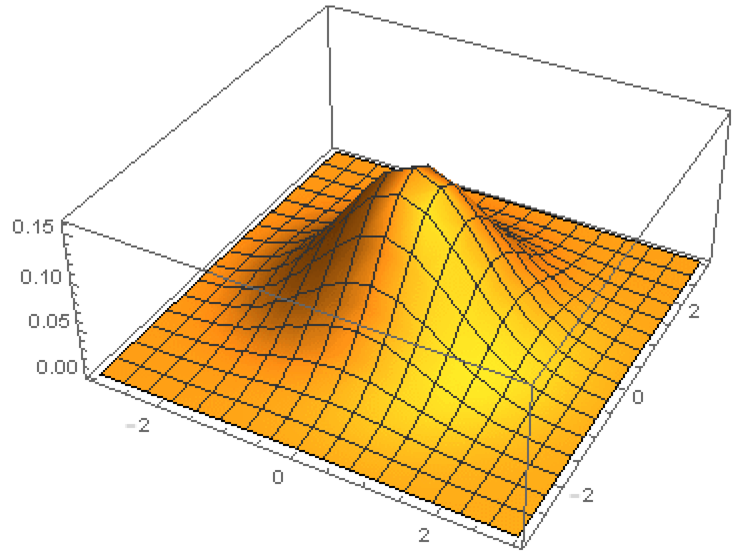
Variance: $\int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$

Estimation of Parameters

- Given samples $x_1, x_2, \ldots x_n$ in R from a probability distribution, estimate the pdf, assuming it is Gaussian
- Mean value: $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$
- Variance: $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$

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

Gaussian pdf in 2D



$$f(x,y) = (2\pi)^{-1} e^{-(x^2+y^2)/2}$$

Bayes Theorem for Parameter Estimation

- Given samples $X = \{x_1, x_2, \ldots x_n\}$ in R from a Gaussian distribution with variance 1, estimate the mean value μ

$$p(\mu|X) = p(X|\mu)p(\mu)/p(X)$$

$p(\mu), p(X)$ are prior pdfs,
$p(X|\mu)$ is the likelihood function for μ
$p(\mu|X)$ is the posterior pdf for μ

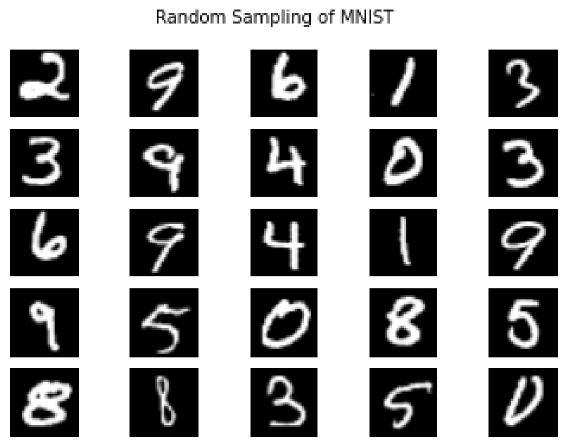
Classification Problem

- Given an image D of a digit, classify it as 0 or 1 or … or 9.

- Let $\theta(i)$ be the hypothesis that the class is i.

- Assume that the probability density functions $p\big(D|\theta(i)\big)$ are known

- The Bayes method gives the best solution



MNIST database and http://andrew.gibiansky.com/blog/machine-learning/k-nearest-neighbors-simplest-machine-learning/

Bayes Solution

$$p(\theta(i)|D) = p\big(D|\theta(i)\big)p\big(\theta(i)\big)/p(D)$$

$p\big(\theta(i)\big)$: prior density
$p(\theta(i)|D)$: posterior density

Find i for which $p(\theta(i)|D)$ is a maximum

The density $p(D)$is unknown, but only the ratios $p(\theta(i)|D)/p(\theta(j)|D)$ are required