Personal Work Diary

Jacob Møller
Aalborg University
Sound Processing

Last compiled: September 28, 2014

Contents

Todo list

Part I

2014

Chapter 1

September

1.1 September 26, 2014

1.1.1 Old Quiz From 2013

These exercises has been answered during summer holiday. The first part is a quiz.

- What is a linear system, and what is meant by its solution?

A system of linear equations is a collection of one or more linear equations involving the same variables. which is not squared or "similar multiplied" (lines or hyper planes). The solution set is all possible solutions which maps the system 'A' to the desired vector 'b'. (x=A) solution if b is in the column space of A (b is linear combination of the columns of Ā).

- What do we mean by consistency of a linear system?

There is either none, one or infinitely many solutions. A system of linear equations is said to be consistent if it has either one solution or infinitely many solutions; a system is inconsistent if it has no solution.

- Explain what is echelon form and reduced echelon form.

Both are triangular form. Echelon Form is when the system is brought to upper triangular form, where all leading entries is the only non-zero in its column. (pivot in all rows except zero-rows.) Reduced Echelon Form is when every leading entry is 1 with zeros above and below.

- Explain the meaning of the following terms in relation to echelon forms: pivot, pivot positions, pivot columns.

pivot: leading entry. pivot position: A pivot position is a location that corresponds to a leading entry in the reduced echelon form of a matrix. pivot column: A pivot column contains a pivot position.

- What is a linear combination of vectors, and what is the relation to Span. . . ?

y = c1v1 + ··· + cpvp is a linear combination, where c is scalars and v is vectors. the spanv1, v2, v3 is a subset of R3 spanned by v1, v2 and v3; That is all linear combinations of v1, v2 qnd v3 . see example 6 pp. 36.

- How can we write a linear system as a matrix equation?

see pp. 33.

- Define independence/dependence in relation to vectors.

if two vectors can be expressed as linear combinations of each other they are linear dependent and vice versa. in a matrix system Ax=0 has only the trivial solution if it is linear independent set. if a set contains more vectors that entries in each vector, the set is linear dependent.

- Explain how we sum, multiply and transpose matrices.

sum : each corresponding element mult: inner product of row and column. tran: inner product of row to col and complex conjugate.

- What do we mean by a matrix inverse?
- When can we invert a matrix? Mention some of the criteria.

1.1.2 Exercise 1 as MATLAB listing

```
% Exercise 4 (Plots)
%
% The hyperbolic paraboloid has the equation y^2/b^2-x^2/a^2=z/c ...
    (c>0), which we will now use.
%
% a) Plot the equation in the xz-plane for a=3, b=2 og c=1.
% b) Put lables, title, etc. on the plot.
% c) Create a contour plot of the function.
% d) Create a surface plot of the function for -20<x<20 og -10<y<10.
% e) Try different viewpoints, colormaps og shadings.
% f) Export one of the figures in a suitable format.
%
%
close all;clear all;
svn_path = '/Users/OSX/Documents/UNI/P7/LaTeX/article_template';
domain = [-20:20];
syms x y;
b = 2; a = 3; c = 1;
f(x,y) = (y.^2./b.^2-x.^2./a.^2)*c;

```

```
% a) Plot the equation in the xz-plane for a=3, b=2 og c=1.
figure
plot(domain,f(0,domain));
xlabel('y-axis');ylabel('f(0,y)');title('Hyperboloid f(x,y) = ...
    (y^2/2^2-x^2/a^2), for x=0');
fig2tikz(gcf,'2d','Graph Plot','Graph Plot',svn_path)

figure
subplot(211)
plot(domain,f(domain,0));
xlabel('y-axis');ylabel('f(x,0)');
subplot(212)
plot(domain,f(0,domain));
xlabel('x-axis');ylabel('f(0,y)');
fig2tikz(gcf,'2d','Comparison of $ f(x,y) $ when $ x=0 $ and $ y=0 ...
    $','Sub Plot',svn_path)

% c) Create a contour plot of the function.
[x y] = meshgrid(-20:1:20,-20:1:20);
figure
contour(x,y,f(x,y))

% d) Create a surface plot of the function for -20<x<20 og -10<y<10.
figure
[x y] = meshgrid(-20:1:20,-10:0.5:10);
f = (y.^2./b.^2-x.^2./a.^2)*c;
surf(x,y,f)
legend('Surf Plot');
ylabel('y-axis'); xlabel('x-axis'); zlabel('f(x,y)');
fig2tikz(gcf,'3d','$ f(x,y) $ in 3 dimensions.','Surf Plot',svn_path)

figure
surfc(x,y,f)
```

1.1.3 Quiz 2

Quiz

1: What do we mean by a transformation T from Rn to Rm, and what is the meaning of its domain, codomain and range?i A transformation (or function or mapping) T from Rn to Rm is a rule that assigns to each vector x in Rn a vector T(x) in Rm. The set Rn is called the domain of T. Rm is called the codomain of T. For x in Rn, the vector T(x) in Rm is called the image of x (under the action of T). The set of all images T (x) is called the range of T .

Quiz

2: What is a linear transformation, and how is it related to a matrix transformation? Every matrix transformation is a linear transformation.

Quiz

3: Regarding mappings, what do we mean by the terms "onto" and "one- to-one"? A mapping$T : Rn \rightarrow Rm$ is said to be onto Rm if each b in Rm is the image of at least one x in Rn. A mapping$T : Rn \rightarrow Rm$ is said to be one-to-one if each b in Rm is the image of at most one x in Rn. (pp. 87)

Quiz

4: Explain the principle behind the LU factorization. Lower and upper trinagular - L:lower - U upper triangular row echelon form of the given matrix.

Quiz

5: What does |A| = 0 imply? A is not invertible.

Quiz

6: How can we find the matrix inverse, using determinants? $A^-1 = 1/det(A) * adj(A)$, where adj(A) is the adjugate of A. (the matrix of co-factors (pp. 204))

Quiz

7: What is the relation between areas/volumes of parallelograms and parallelepipeds and determinants? If A is a 2×2 matrix, the area of the parallelogram determined by the columns of A is |det A|. If A is a 3×3 matrix, the volume of the parallelepiped determined by the columns of A is |det A|.

Quiz

8: Answer Q7, for linear transformations of parallelograms and parallelepipeds.

1.1.4 Exercise 2 as MATLAB listing

```
%
% Exercise 1: Let A = [0 2;0 2] and define T : R2 to R2 by T(x) = Ax.
% Find the images under T of u=[1 -3]T and v=[ab]T.
A = eye(2)*2
u = [1 -3 ]'
```

```
x = A\u
x = ref([A u])

display('The same applies for vector v. x = [a/2 b/2]T')
pause;

% Let T (x) = Ax, and
% Find a vector x whose image under T is b, and determine whether x is ...
    unique.
A = [1 0 -2 ; -2 1 6 ; 3 -2 -5 ]
b = [-1 -7 -3 ]'
Ab_red = ref([A b])
x = A\b
display('x is unique which is seen from the reduced Ab matrix, since ...
    it has full rank.')
pause;

% Solve Ax = b using LU factorization and by ordinary row reduction, with
A = [4 3 -5 ; -4 -5 7 ; 8 6 -8 ]
b = [2 -4 6 ]'
Ab_red = ref([A b])
[L,U] = ref_lu(A)
Lb_red = ref([L b])
y = Lb_red(:,end)
Uy_red = ref([U y])
x = Uy_red(:,end)
pause;

% Compute determinants using cofactor expansion both across the first ...
    row and down the second column.
A = [3 0 4 ; 2 3 2 ; 0 5 -1 ]
detA = det(A)
detA = (-1)^(1+1) * 3*(3*(-1)-5*2) + (-1)^(1+3) * 2*(5*4-0)
detA = (-1)^(2+2) * 3*((-1)*3-0*4) + (-1)^(3+2) * 5*(2*3-4*2)

B = [ 2 -4  3 ; ...
      3  1  2 ; ...
      1  4 -1 ]
detB =  det(B)
detB = (-1)^(1+1) * 2  * (1*(-1)-4*2)   + (-1)^(1+2) * (-4) * ...
    ((-1)*3-2*1)   + (-1)^(1+3) * 3 *(3*4-1*1)
detB = (-1)^(1+2) *(-4)* (3*(-1)-(1*2)) + (-1)^(2+2) *    1 *( ...
    2*(-1)-3*1)   + (-1)^(3+2) * 4 *(2*2-3*3)

% Compute detA by row reduction to echelon form.
A = [1 5 -6 ; -1 -4 4 ; -2 -7 9 ]
detA = det(A)
[p,L,U] = ref_plu(A)
diagU = prod(diag(U))
pause;
%  Use determinants to check if A is invertible.
```

```
A = [2 3 0 ; 1 2 1 ; 1 2 1 ]
detA = det(A)
display('Since determinant != 0, A is invertible')
pause;

% Use Cramer's rule to compute the solutions of
%    3x1 - 2x2 =  7
%   -5x1 + 6x2 = -5
%% and
%   -5x1 + 3x2 = 9
%    3x1 - x2 = -5.
A = [5 -3 ; ...
     3 -2 ]
b = [9 -5]'
A1b=A; A1b(:,1) = b
A2b=A; A2b(:,2) = b
x1 = det(A1b)/det(A)
x2 = det(A2b)/det(A)
```

1.1.5 Quiz 3

Quiz

1: What is eigenvalues and -vectors? An eigenvector of an n×n matrix A is a nonzero vector x such that $Ax = \lambda x$ for some scalar λ. A scalar λ is called an eigenvalue of A if there is a nontrivial solution x of $Ax = \lambda x$; such an x is called an eigenvector corresponding to λ. (pp.303)

Quiz

2: How can we find the eigenvalues of a triangular matrix? The eigenvalues of a triangular matrix are the entries on its main diagonal.

Quiz

3: What is the characteristic equation? Useful information about the eigenvalues of a square matrix A is encoded in $A - \lambda I$. This matrix fails to be invertible precisely when its determinant is zero. So the eigenvalues of A are the solutions of the equation $det(A - \lambda I) = 0$, which is the characteristic equation.

Quiz

4: What does it mean that a matrix is diagonalizable, and how can such a matrix be diagonalized? A square matrix A is said to be diagonalizable if A is similar to a diagonal matrix, that is, if $A = PDP^{-1}$ for some invertible matrix P and some diagonal matrix D. $A^3 = PD^3 * P^{-1}$. D has the eigenvalues on the diagonal. P contains the eigenspace which spans Rn.

Quiz

5: Mention a condition for a matrix to be diagonalizable? An n×n matrix A is diagonalizable if and only if A has n linearly independent eigenvectors. In fact,$A = PDP - 1$, with D a diagonal matrix, if and only if the columns of P are n linearly independent eigenvectors of A. In this case, the diagonal entries of D are eigenvalues of A that correspond, respectively, to the eigenvectors in P. In other words, A is diagonalizable if and only if there are enough eigenvectors to form a basis of Rn. We call such a basis an eigenvector basis.

Quiz

6: What are the princples behind the (inverse) power methods? $A^k = P * D^k * P^-1$.

Quiz

7: Explain the following terms in relation to vectors: inner product, length, orthogonal, distance, projection.

Quiz

8: What is the characteristics of orthogonal and orthonormal vector sets?

Quiz

9: What is the general least squares problem? Mention some real-life examples of least squares problems.

1.1.6 MSPR - First Exercises with MGC

Tag(s):
Courses in General (courses)
Pattern Recognition Course (mspr)

Author(s):
Jacob Møller (jm)

Quiz mm1

The quiz is answered in the following.

- What is Classification?
- What is a Prior?
- What quantities must be known for the quadratic and linear classifiers to work, based on the multivariate normal distribution?
- Under what condition is it possible to classify classes with the same mean?
- Do the decision boundaries for the quadratic and linear classifiers look alike?

Classification

Classification is the act of assigning a class label to an object, a physical process or an event. For automated sorting we have to classify the objects, by measuring some properties of the individual objects.
Measurement vectors are processed to reveal information for the task. Models are made by describing the object. In some cases it is not trivial to define relevant classes. It is like sorting parameters in to classes. License Plate characters are easy. Sorting tomatoes into three different classes are not so easy. For sorting electronic parts, the classes are "IC's", "resistors" and so on.

Detection - A Special Case of Classification

Here, only two class labels are available; yes and no.

Identification - A Special Case of Classification

e.g. Fingerprint recognition or face recognition. Usually by using a large database.

1.1.7 How to Design a pattern classifier

Bayesian Theoretic Framework is a solid base for pattern classification. We have both a prior and a posterior. The starting point is a stochastic experiment defined by a set $\Omega = \{\omega_1, \cdots \omega_k\}$. We assume the classes are mutually exclusive.

Prior and Posterior

Prior

The probability $P(\omega_k)$ of having a class ω_k is called the *prior probability*. It represents the knowledge we have about the object before measurements. With K different classes that is,

$$\sum_{1}^{K} P(\omega_k) = 1 \tag{1.1}$$

We produce a measurement vector called $\mathbf{z}$, with dimension N. These vary, even within the same class, due to noise and i.e. eccentricities of bolts are not fixed size and randomness due to other noise. Therefore, we have the *conditional probability function* of $\mathbf{z}$, denoted as $p(\mathbf{z}|\omega_k)$.**lecture words**: *"Specific distribution of of the data* $\mathbf{z}$ *for one class* ω_k*"*.

It is the density of $\mathbf{z}$ coming from a known class ω_k
If $\mathbf{z}$ comes from an object with an unknown class, it is denoted by $p(\mathbf{z})$.

$$p(\mathbf{z}) = \sum_{1}^{K} p(\mathbf{z}|\omega_k)P(\omega_k) \tag{1.2}$$

lecture words: *"Collecting all possible measurements and plotting a scatter plot"* Due to the mutually exclusive of ω_k, $p(\mathbf{z})$ is unconditional.

The Decision Function

The pattern classifier casts the measurement vector in the class that will be assigned to the object. This is done by the decision function. The decision function $\hat{\omega}(.)$ maps the measurements space onto the set of possible classes. That is, for $\mathbf{z}$ N dimensional, $\mathbb{R}^N \to \Omega$

Bayes Classifier

An Erroneous assignment of a class to an object leads to an impairment of its usefulness. A *Bayes Classifier* is a pattern classifier that is based on the following two:

- The loss of value, when an object is erroneously classified, can be quantified as a cost.
- The expectation of the cost is acceptable for optimization criterion.

By meeting these two, the development of a optimal pattern classification is straight forward, if you have good estimates of the densities of the classes.

The Cost Function

The loss of value is quantified by a *Cost Function* $C(\hat{\omega}|\omega_k)$. The function $C(.|.)$: Ω x $\Omega \to \mathbb{R}$ Expressing the cost is when the class $\hat{\omega}$ is assigned to an object when the true class is ω_k.

Posterior Probability and Minmum Risk Classification

Bayes Theorem of conditional probabilities says,

$$P(\omega_k|\mathbf{z}) = \frac{p(\mathbf{z}|\omega_k)P(\omega_k)}{p(\mathbf{z})} \tag{1.3}$$

When we assign $\hat{\omega}$ to $\mathbf{z}$, from the object with true class ω_k, the cost $C(\hat{\omega}|\omega_k)$ is involved. The posterior prob. of having such an object is: $P(\hat{\omega}|\mathbf{z})$. Therefore, the expectation of the cost is, the conditional risk:

$$R(\hat{\omega}) = \mathbb{E}[C(\hat{\omega}|\omega_k)|\mathbf{z}] = \sum_{k=1}^{K} C(\hat{\omega}|\omega_k)P(\hat{\omega}|\mathbf{z}) \tag{1.4}$$

Which gives us the overall risk, over the full measurement space:

$$R = \mathbb{E}[R(\hat{\omega}(\mathbf{z})|\mathbf{z})] = \int_{\mathbf{z}} R(\hat{\omega}(\mathbf{z})|\mathbf{z})p(\mathbf{z})d\mathbf{z} \tag{1.5}$$

This is the size we want to minimize:

$$\hat{\omega}_{\mathtt{BAYES}}(\mathbf{x}) = \underset{\omega \in \Omega}{\operatorname{argmin}} \{R(\omega|\mathbf{z})\} \tag{1.6}$$

Which is described for $\mathbf{z}$ as,

$$\begin{aligned}
\hat{\omega}_{\mathtt{BAYES}}(\mathbf{z}) &= \underset{\omega \in \Omega}{\operatorname{argmin}} \left\{ \sum_{k=1}^{K} C(\omega|\omega_k) P(\omega_k|\mathbf{z}) \right\} \\
&= \underset{\omega \in \Omega}{\operatorname{argmin}} \left\{ \sum_{k=1}^{K} C(\omega|\omega_k) \frac{p(\mathbf{z}|\omega_k)P(\omega_k)}{p(\mathbf{z})} \right\} \\
&= \underset{\omega \in \Omega}{\operatorname{argmin}} \left\{ \sum_{k=1}^{K} C(\omega|\omega_k) p(\mathbf{z}|\omega_k) P(\omega_k) \right\}
\end{aligned}$$

This is minimum risk classification or Bayesian classification.

1.1.8 Normal distributed Measurements, linear and quadratic classifiers

- What quantities must be known for the quadratic and linear classifiers to work, based on the multivariate normal distribution?

Since the conditional probability distributions are modelled as normal, the quantities of the expectation vector $\boldsymbol{\mu}_k$ and covariance matrix $\mathbf{C}_k$, must be known, see pp. 26.
A quadratic classifier has a quadratic decision function. The boundaries are then pieces of quadratic hyper-surfaces. The boundaires are expressed in eq(2.23) pp. 27.

1.1.9 Class inpendent covarince matrices

If you assume no correlation between the classes - covariance matrices do not depend on the classes i.e. $\mathbf{C}_k = \mathbf{C}$ for all $\omega_k \in \Omega$. This is when $\mathbf{z} = \boldsymbol{\mu}_k + \mathbf{n}$. Then the decision function is linear see eq (2.24) pp 28.

1.1.10 Class inpendent expectation vectors

- Under what condition is it possible to classify classes with the same mean?

When the expectation vectors do not depend on the class: $\boldsymbol{\mu}_k = \boldsymbol{\mu}$, for all k. The central parts of the conditional pdf, overlap. The decsion can be seen in eq (2.29) pp.31. I belive this can be used with the nmc() function in PRTools (nearest mean classifier).

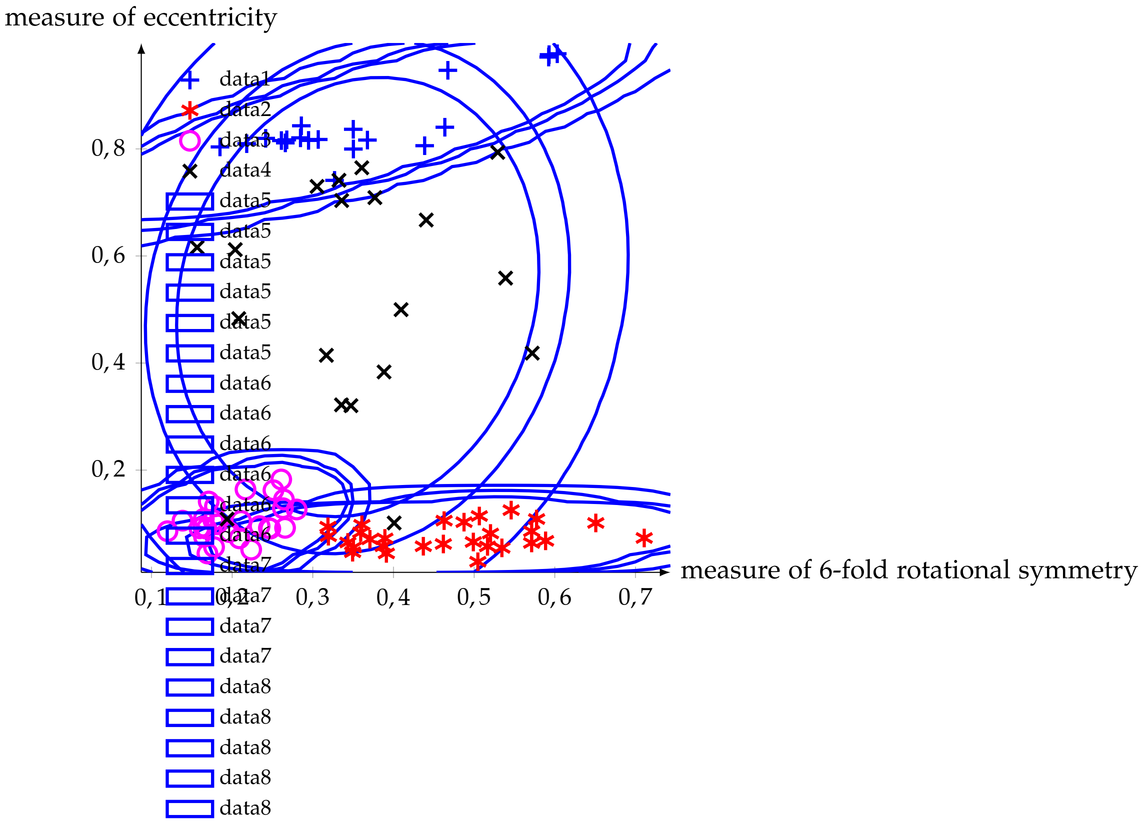

Figure 1.1: probability densities of the measurements

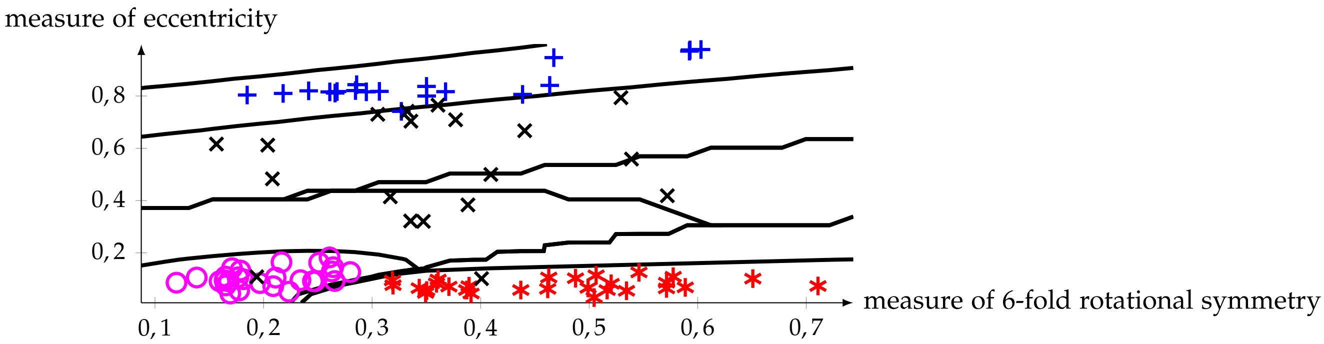

Figure 1.2: Estimating Decision Boundaries with inclusion of function

Do exercise 2.4 and 2.5

Insert the subplot of bayes_classification_21_22 and describe it

understand gaussm()

Tag(s):
Pattern Recognition Course (mspr)
Courses in General (courses)
Exercises (exercises)

Author(s):
Jacob Møller (jm)

Author Index

Tag Index