Lecture 19: Data fitting

Last time we have derived the important formula

$$P = A(A^T A)^{-1} A^T .$$

which gives the projection matrix when projecting onto the image of a matrix A.

Given a system of linear equations $Ax = b$, the point $x = (A^T A)^{-1} A^T b$ is called the **least square solution** of the system.

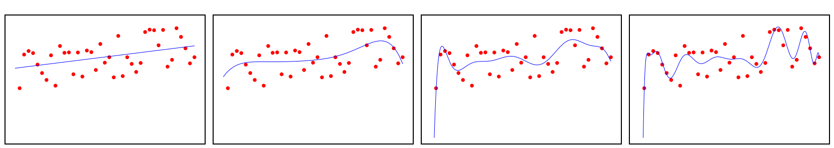
If A has no kernel, then the least square solution exists.

Proof. We know that if A has no kernel then the square matrix $A^T A$ has no kernel and is therefore invertible.

In applications we do not have to worry about this. In general, A is a $n \times m$ matrix where n is much larger than m meaning that we have **lots** of equations but few variables. Such matrices in general have a trivial kernel. For linear regression for example, it only appears if all data points are on a vertical axes like $(0,2), (0,6), (0,0), (0,4)$ and where any line $y = mx + 3$ is a least square solution. If in real life you should get into a situation where A has a kernel, you use the wrong model or have not enough few data.

If x is the least square solution of $Ax = b$ then Ax is the closest point on the image of A to b. The least square solution is the best solution of $Ax = b$ we can find. Since $Px = Ax$, it is the closest point to b on V. Our knowledge about kernel and the image of linear transformations helped us to derive this.

1 Finding the best polynomial which passes through a set of points is a **data fitting** problem. If we wanted to accommodate **all data**, the degree of the polynomial would become too large. The fit would look too wiggly. Taking a smaller degree polynomial will not only be more convenient but also give a better picture. Especially important is **regression**, the fitting of data with linear polynomials.

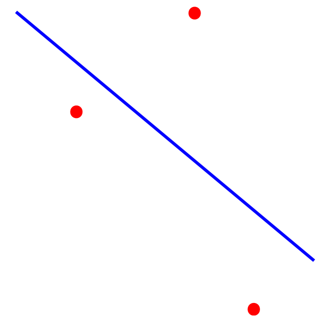


The above pictures show 30 data points which are fitted best with polynomials of degree 1, 6, 11 and 16. The first linear fit maybe tells most about the trend of the data.

2 The simplest fitting problem is fitting by lines. This is called linear regression. Find the best line $y = ax + b$ which fits the data

x	y
-1	1
1	2
2	-1

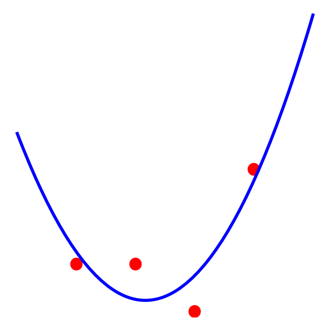
Solution. We will do this in class. The best solution is $y = -x/2 + 1$.



3 Find the best parabola $y = ax^2 + bx + c$ which fits the points

x	y
-1	8
0	8
1	4
2	16

We do this in class. The best solution is $f(x) = 3x^2 - x + 5$.



4 Find the function $y = f(x) = a\cos(\pi x) + b\sin(\pi x)$, which best fits the data

x	y
0	1
1/2	3
1	7

Solution: We have to find the least square solution to the system of equations

$$\begin{aligned} 1a + 0b &= 1 \\ 0a + 1b &= 3 \\ -1a + 0b &= 7 \end{aligned}$$

which is in matrix form written as $A\vec{x} = \vec{b}$ with

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \end{bmatrix}, \vec{b} = \begin{bmatrix} 1 \\ 3 \\ 7 \end{bmatrix} .$$

Now $A^T\vec{b} = \begin{bmatrix} -6 \\ 3 \end{bmatrix}$ and $A^TA = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ and $(A^TA)^{-1} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1 \end{bmatrix}$ and $(A^TA)^{-1}A^T\vec{b}$ is $\begin{bmatrix} -3 \\ 3 \end{bmatrix}$. The best fit is the function $\boxed{f(x) = -3\cos(\pi x) + 3\sin(\pi x)}$.

5 Find the circle $a(x^2 + y^2) + b(x + y) = 1$ which best fits the data

x	y
0	1
-1	0
1	-1
1	1

In other words, find the least square solution for the system of equations for the unknowns a, b which aims to have all 4 data points (x_i, y_i) on the circle. To get system of linear equations $Ax = b$, plug in the data

$$\begin{aligned} lla + b &= 1 \\ a - b &= 1 \\ 2a &= 1 \\ 2a + 2b &= 1 . \end{aligned}$$

This can be written as $Ax = b$, where

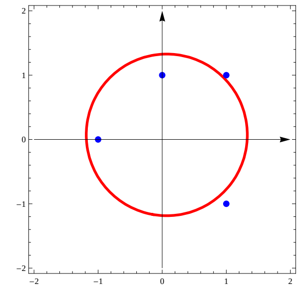
$$A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 2 & 0 \\ 2 & 2 \end{bmatrix}, \ b = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} .$$

We get the least square solution with the usual formula. First compute

$$(A^TA)^{-1} = \begin{bmatrix} 3 & -2 \\ -2 & 5 \end{bmatrix} /22$$

and then

$$A^Tb = \begin{bmatrix} 6 \\ 2 \end{bmatrix} ,$$



Homework due March 23, 2011

1 Find the function $y = f(x) = ax^2 + bx^3$, which best fits the data

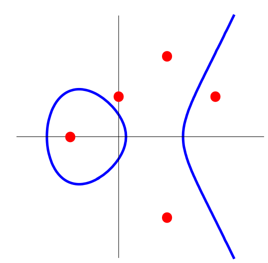
x	y
-1	1
1	3
0	10

2 A curve of the form

$$y^2 = x^3 + ax + b$$

is called an **elliptic curve** in Weierstrass form. Elliptic curves are important in cryptography. Use data fitting to find the best parameters (a, b) for an elliptic curve given the following points:

$$\begin{aligned} (x_1, y_1) &= (1, 2) \\ (x_2, y_2) &= (-1, 0) \\ (x_3, y_3) &= (2, 1) \\ (x_4, y_4) &= (0, 1) \end{aligned}$$



3 Find the function of the form

$$f(t) = a\sin(t) + b\cos(t) + c$$

which best fits the data points $(0, 0), (\pi, 1), (\pi/2, 2), (-\pi, 3)$.