[I] Linear regression: Details

let $x_1, x_2 \dots x_m$ be given numbers and let $y_1 \dots y_m$ be some associated values.

Find a polynomial $p(x) = a_0 + a_1 x$ such that it fits to the given data 'as exactly as possible'

$\Rightarrow$ the exact fitting $\rightarrow$ $a_0 + a_1 x_k = y_k$ for all $(k = 1, 2 \dots M)$

$$\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \Rightarrow \text{No exact solution.}$$

So best fitting is by minimizing the norm of the residual

$$F(a) := \| Aa - y \|^2 \longrightarrow \min$$

$$\Rightarrow F(a_0, a_1) := \sum_{k=1}^{M} (a_0 + a_1 x_k - y_k)^2 \longrightarrow \min$$

~~At~~ At the minimum, the partial derivates vanish:

$$\frac{\partial F}{\partial a_0} = \sum_{k=1}^{M} 2 (a_0 + a_1 x_k - y_k) \cdot 1 = 0$$

$$\frac{\partial F}{\partial a_1} = \sum_{k=1}^{M} 2 \cdot (a_0 + a_1 x_k - y_k) \cdot x_k = 0$$

$\Rightarrow a$ & a_1 can be computed by solving the system of equations:

$$\sum_{k=1}^{M} y_k = a_0 \sum_{k=1}^{M} 1 + a_1 \sum_{k=1}^{M} x_k$$

$$\sum_{k=1}^{M} x_k y_k = a_0 \sum_{k=1}^{M} x_k + a_1 \sum_{k=1}^{M} x_k^2$$

By direct calculation $\Rightarrow$ $A^* A a = A^* y$

(2) Quadratic reg: Details.

let $x_1, \dots x_M$ be given ~~values~~ Numbers and let $y_1 \dots y_M$ be some associated Values ($M \geq 3$)
Find polynomial $p(x) = a_0 + a_1 x + a_2 x^2$ such that it fits to the given data "as exactly as possible"

The exact fitting: $a_0 + a_1 x_1 + a_2 x_1^2 = y_1$

$$\vdots$$

$$a_0 + a_1 x_M + a_2 x_M^2 = y_M$$

There is no exact sol in general. It is worth defining the 'best fitting' by minimizing the norm of residual:

$$F(a) := \| Aa - y \|^2 \to \min!$$

$$F(a_0, a_1, a_2) = \| A\underline{a} - \underline{y} \|^2 = \sum_{K=1}^{M} (a_0 + a_1 x_K + a_2 x_K^2 - y_K)^2 \to \min!$$

$$\frac{\partial F}{\partial a_0} = \sum_{K=1}^{M} 2 (a_0 + a_1 x_K + a_2 x_K^2 - y_K) \cdot 1 = 0$$

$$\frac{\partial F}{\partial a_1} = \sum_{K=1}^{M} 2 (a_0 + a_1 x_K + a_2 x_K^2 - y_K) \, x_K = 0$$

$$\frac{\partial F}{\partial a_2} = \sum_{K=1}^{M} 2 (a_0 + a_1 x_K + a_2 x_K^2 - y_K) \cdot x_K^2 = 0$$

whence a_0, a_1 can be computed. the coefficients a_q

$$a_0 \cdot \left(\sum_{K=1}^{M} 1\right) + a_1 \left(\sum_{K=1}^{M} x_K\right) + a_2 \sum_{K=1}^{M} (x_K^2) = \sum_{K=1}^{M} y_K$$

$$a_0 \cdot \left(\sum_{K=1}^{M} x_K\right) + a_1 \left(\sum_{K=1}^{M} x_K^2\right) + a_2 \left(\sum_{K=1}^{M} x_K^3\right) = \sum_{K=1}^{M} y_K \, x_K$$

$$a_0 \cdot \left(\sum_{K=1}^{M} x_K^2\right) + a_1 \left(\sum_{K=1}^{M} x_K^3\right) + a_2 \left(\sum_{K=2}^{M} x_K^4\right) = \sum_{K=1}^{M} y_K \, x_K^2$$

Solutions of linear Systems of equations in the Sense of least squares,

Let $A \in M_{M \times N}$ be a Matrix $b \in \mathbb{R}^M$. the Vector

$x^+ := A^+ b \in \mathbb{R}^N$ → The generalized Solutions of the system $Ax = b$

There is examples in lect 6

The Jacobi iteration

Let $A \in M_{n \times n}$ be a regular Matrix, Let $b \in \mathbb{R}^N$ be a vector and Consider the system of equations

$$Ax = b$$

1) Theorem:
The inverse of a lower triangular matrix is also a lower triangular matrix, and the inverse of a upper triangular matrix is also an upper triangular matrix.
if the matrix A is a (lower/upper/diagonal) matrix so the diagonal elements of A^{-1} are the reciprocal values of A diagonal elements.
Aim: Find the inverse of the Matrices when we have triangular nonsingular matrix

2) Aim: find the eigen values and the eigen vectors of a square matrix:
·Eigen values: it involves solving the characteristic equation $\det(A - I\lambda) = 0$ wher A is the matrix and λ is the eigenvalue and I is the Identity matrix.
Eigen vectors: after finding the eigenvalues, we can find the eigenvectors by solving the system of linear equations $(A - \lambda I)V = 0$ for each eigenvalue.

3) Aim: the Gauss and Gauss-jordan elemenation methods aim to find the solution of linear equations system by elemination using row operations to transform the augmented matrix.
Theorem: Gauss Elemination: involves row operation to transform the augmented matrix into upper tringular form then solving the new system of equations using the back substitution
Gauss-Jordan Elemination: extends the Gauss elimination to reduce the matrix to the Identity form directly yeilding the solution vector
Assumption: the coefficient matrix is nonsingular and the eqations has a unique solution

6) Aim: Gram-Schmidt Orthogonalization aims for transforming a set of linearly independent vectors into a set of orthogonal vectors

Theorem: let $a_1, a_2, a_3, \ldots a_n \in \mathbb{R}^N$ be linealy independent vectors, we can define:

$$\tilde{e}_k := a_k - \sum_{j=1}^{k-1} \langle a_k, e_j \rangle \cdot e_j \iff e_k = \frac{\tilde{e}_k}{\|\tilde{e}_k\|}$$

wher $\tilde{e}_1 = a_1 \iff e_1 = \frac{\tilde{e}_1}{\|\tilde{e}_1\|}$ & $k = 2, 3, \ldots, N$

and the vectors $e_1, e_2, \ldots, e_N$ form an orthonormal vector system.

Note: $\langle e_k, e_j \rangle = 0$ if $k \neq j$ and $\|e_k\| = 1$.

4) Aim: LU Decompostion decomposes the matrix into the product of L(lower triangular) and U(upper triangular) matrices.

Theorem: Let ~~A~~ $A = [a_{kj}] \in M_{N \times N}$ a regular matrix, for which the Gaussian Elemination can be performed without swapping rows. The matrix A can be uniquely decomposed in the following form: $A = L \cdot U$; $\det A = \det L \cdot \det U = u_{11} \cdot u_{22} \cdot u_{33} \cdots u_{NN}$

where L is normal lower triangular matrix ($L_{kk} = 1$ & $L_{kj} = 0$ if $j > k$) and U is upper triangular matrix ($u_{kj} = 0$ if $j < k$)

~~linear~~ linear system $Ax = b$ should be solved

$Ly = b$ & $Ux = y \Rightarrow$ so only forward and backward ~~substitu~~ substitutions have to be performed.

5) LDL^* and Cholesky Decomposition aim to factorize a symmetric positive definite matrix into a product of a lower triangular matrix L, a diagonal matrix D and the transpose of L.

Theorem: Let $A \in M_{N\times N}$ self-adjoint, positive definite matrix (which assures that the gaussian Elemination can be performed without swapping rows)

Considre: $A = LU$. and let D be the diagonal part of U then $U = DU'$ where U' is a normed ~~diagonal~~ upper triangular matrix $\Rightarrow$ A self-adjoint and withe the LU decomposition implies:

$$A = LU = LDU' \; ; \; U' = L^* \Rightarrow A = LDL^*$$

Theorem: Let $A \in M_{N\times N}$ be self-adjoint, positive definite matrix with the LDL^* decomposition:

$$A = L_0 D L_0^*$$

$$\Rightarrow A = (L_0\sqrt{D})(L_0\sqrt{D})^* = LL^*$$

where L is a not necessarily normed lower triangular matrix, with positive diagonal elements.

system $Ax = b$ should be solved $Ly = b$, $L^*x = y \Rightarrow$ only forward & backward substitutions have to be performed

7) QR decomposition: Aims to factorize the matrix into a product of Q (orthogonal matrix) and R (upper triangular matrix)

Theorem: Q is calculated throw Gram-Schmidt process from A

$$Q = \left(e_1 \middle| e_2 \middle| \cdots \middle| e_N\right) \xLeftarrow[\text{Schmidt}]{\text{Gram}} A = \left(a_1 \middle| a_2 \middle| \cdots \middle| a_N\right)$$

While R is:

Note $r_{11} = \|\tilde{e}_1\|$
$r_{22} = \|\tilde{e}_2\|$ ——

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} & \cdots \\ 0 & r_{22} & r_{23} & \cdots \\ \vdots & \vdots & \ddots & \\ & & & r_{NN} \end{pmatrix} \quad \text{where } r_{kj} = \langle e_k, a_j \rangle$$

system $Ax = b$ can be solved: $Rx = Q^{-1}b = Q^{*}b$

8) SVD: Singular Value Decomposition aims to decompose the matrix A into a product of thee matrices $A = USV^{*}$ where this decomposed is NOT uniquely and A is NOT necessarly square

Theorem of A square (regular): Let $A \in M_{N,N}$, Denoted by $\lambda_j > 0$ the eigen values of the (self-adjoint, positive definite matrix ~~A*~~ $A^{*}A$. and v_j are the corrisponding orthonormal eigenvectors ($j = 1, 2, \ldots, N$).

$V = \left(V_1 \middle| V_2 \middle| \cdots \middle| V_N\right)$ from the singular values: $\sigma_k = \sqrt{\lambda_k}$ ($k = 1, 2, \ldots, N$)

$$\Rightarrow S = \begin{pmatrix} \sigma_1 & 0 & 0 & \cdots \\ 0 & \sigma_2 & 0 & \cdots \\ \vdots & & \ddots & \\ & & & \sigma_N \end{pmatrix}$$

we can compute $u_k = \frac{A v_k}{\sigma_k}$ $\Rightarrow U = \left(u_1 \middle| u_2 \middle| \cdots \middle| u_N\right)$

9) scalar product (Inner product): Let $(x_1, x_2, \dots x_N) \in$ & $y = (y_1, y_2, \dots y_N) \in \mathbb{R}^N$ be N-dimensional vectors.

we define the scalar products as:

$$\langle x, y \rangle := \sum_{k=1}^{N} x_k y_k$$

we use the scalar product to generat the Euclidean norm as following: $\|x\| = \sqrt{\langle x, x \rangle}$

properties: $\langle x, x \rangle \geq 0$ and $= 0$ only if $x = (0, 0, \dots, 0)$

$\langle x, y \rangle = \langle y, x \rangle$ $\quad \langle \alpha x, y \rangle = \alpha \langle x, y \rangle$

$\langle x+y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$

~~useful~~

vector norms: Let $x := (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$ be an N-dimensional vector then:

Max norm $\|x\|_{max} := \max_{1 \leq k \leq N} |x_k|$

sum norm $\|x\|_1 := \sum_{k=1}^{N} |x_k|$

Euclidean norm $\|x\|_2 = \sqrt{\sum_{k=1}^{N} |x_k|^2}$

Matrix norm: let $A \in M_{N \times M}$ matrix of type $N \times M$ then:

Max norm $\|A\| := \max\{|A_{kj}| : 1 \leq k \leq N \ \& \ 1 \leq j \leq M\}$

sum norm $\|A\| := \sum_{k=1}^{N} \sum_{j=1}^{M} |A_{kj}|$

Forbenius norm $\|A\| := \sqrt{\sum_{k=1}^{N} \sum_{j=1}^{M} |A_{kj}|^2}$

matrix norm induced by a vector norm is much more important

$\|A\| := \max\{\|Ax\| : x \in \mathbb{R}^M, \|x\| \leq 1\}$ wher $\|Ax\| \leq \|A\| \cdot \|x\|$ for every $x \in \mathbb{R}^M$

Row norm $\|A\| = \max_{1 \leq k \leq N} \sum_{j=1}^{M} |A_{kj}|$

column norm $\|A\| = \max_{1 \leq j \leq M} \sum_{k=1}^{N} |A_{kj}|$

Euclidean norm $\|A\| = \max |\lambda_k|$; A self-Adjoint

$\boxed{\text{self-adjoint positive definit Euclidean norm } \|A\| = \max_{1 \leq k \leq N} \sqrt{\lambda_k}}$

fixed point theorem! Suppose that the mapping $F: \mathbb{R}^N \longrightarrow \mathbb{R}^N$ is contraction, there exists a constant $0 \leq q \leq 1$ such that for arbitary $x, y \in \mathbb{R}^N$, the followin inequality holds:

$$\|F(x) - F(y)\| \leq q \cdot \|x - y\|$$ the the equation $x = F(x)$ has a unique solution $x^* \in \mathbb{R}^N$, which is called the fixed point of F.

for arbitary vector $x_0 \in \mathbb{R}^N$, the vector sequence

$$x^{(n+1)} := F(x^{(n)}) \; ; n = 0, 1, -$$

converges to the fixed point x^*.

the error of nth element:

$$\|x^{(n)} - x^*\| \leq q^n \cdot \|x^{(0)} - x^*\|.$$

~~Let B ∈ M~~

Let $B \in M_{N \times N}$ given matrix and $f \in \mathbb{R}^N$ given vector so the linear systems: $x = Bx + f$

if any of the matrix norms (Row norm, column norm) is less than 1 the the system has a unique solution x^*. for a starting vector $x^{(0)}$ the vector sequence defined by:

$$x^{(n+1)} := Bx^{(n)} + f \; ; n = 0, 1, 2 --$$

where $\|x^{(n)} - x^*\| \leq \|B\|^n \cdot \|x^{(0)} - x^*\|$ is the error of the nth term.

and another ~~post~~ priori error estimation

$$\|x^{(n)} - x^*\| \leq \frac{\|B\|^n}{1 - \|B\|} \cdot \|x^{(1)} - x^{(0)}\|$$

11) we aim to estimate the number of iterations required for ~~...~~ achiving an apropriate error.
for a specific error $\|x^{(n)} - x^*\| = e^*$
~~there is~~ wer write $e^* \leq \frac{\|B\|^n}{1-\|B\|} \|x^{(1)} - x^{(0)}\|$
so we calculat $x^{(1)}$ from the given $x^{(0)}$
and we use the last equation to find n from
as $n \geq \frac{\log\left(\frac{e^*}{\|x^{(1)}-x^{(0)}\|}(1-\|B\|)\right)}{\log(\|B\|)}$

12) The Richardson iteration is an iterative method for solving linear systems.
Let $A \in M_{N\times N}$ be self-adjoint, positive definite matrix. for each parameter $0 \leq \omega < \frac{2}{\|A\|}$ (wher $\|A\|$ is an arbitary matrix norm induced by vector norm):
$\rho(I-\omega A) < 1$ $\boxed{\rho(A) := \max_{1\leq k\leq N} |\lambda_k|}$ is the spectral radius and $\rho(A) \leq \|A\|$ always
Consequently, the Richardson iteration is convergent.
we can define $\omega := \frac{1}{\|A\|}$ but we should define ω in optimal way to get the iteration as fast as possible.
Thespectral radius $\rho(I-\omega A)$ is the least $\Rightarrow$ the convergence is the fastest), when $\omega = \frac{2}{\lambda_{max} + \lambda_{min}}$
$\Rightarrow \rho(I-\omega A) = \frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}}$
if the quotient $\lambda_{max}/\lambda_{min}$ (the condition number of A) is great $\Rightarrow$ the convergence remains slow.

(14) Gershgorin's Circles provide a method To approximate the location of eigenvalues of a square matrix:

Theorem: the eigenvalues lie within the union of circles in the complex plane, the discks are centered at the diagonal elements of the matrix, with a radii given by the sum of the absolute values of the non-diagonal elements of the matrix.

we use this method to prove that 0 is or isn't an possible eigenvalue so if 0 isn't ~~expossible~~ included within the circles $\Rightarrow$ A is regular

(15) The power iteration: aims to find dominant eigenvalues and its associated eigenvectors of a matrix

Theorem: Let $A \in M_{N \times N}$ be a normal matrix with the eigenvalues $|\lambda_1| \leq |\lambda_2| \leq \dots \leq |\lambda_{N-1}| < |\lambda_N|$, and $s_1, s_2, \dots, s_N$ are the orthonormal eigenvectors.

Let $x_0 \in \mathbb{R}^N$ be a vector such that $\langle x_0, s_N \rangle \neq 0$

and define: $x_{n+1} := A x_n \; ; n = 0, 1, \dots$

Then the sequance of Rayleigh quotients

$$\frac{\langle A x_n, x_n \rangle}{\|x_n\|^2} = \frac{\langle x_{n+1}, x_n \rangle}{\|x_n\|^2}$$ converges to λ_N

for the speed of convergence we have the followings estimation: $$\left| \frac{\langle A x_n, x_n \rangle}{\|x_n\|^2} - \lambda_N \right| \leq \text{const.} \left| \frac{\lambda_{N-1}}{\lambda_N} \right|^{2n}$$

16) Shifted inverse power method: Aims to find approximating eigenvalues and eigen vectors of a square matrix with the option of introducing a shift for better convergence.

Theorem: Let $A \in M_{N\times N}$ be a normal matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_N$ and orthonormal eigenvectors $s_1, s_2, \ldots, s_N$.

assume that λ_k is a single eigenvalue and λ_0 is a sufficiently good approximation of λ_k; such that λ_k is the nearest eigenvalue to λ_0; when for every $j \neq k$ $|\lambda_j - \lambda_0| > |\lambda_k - \lambda_0|$ holds.

Then the number $\lambda_k - \lambda_0$ is an eigenvalue of the matrix $(A - \lambda_0 I)$ with minimal absolute value, so that the inverse iteration is applicable.

~~So the inverse shifted iteration~~

So the shifted inverse iteration Theorem said,

Let $x_0 \in \mathbb{R}^N$ be a vector such that $\langle x_0, s_k \rangle \neq 0$ and define: $x_{n+1} := (A - \lambda_0 I)^{-1} x_n$, $n = 0, 1, 2, \ldots$

Then the sequence of Rayleigh quotients $\frac{\langle x_{n+1}, x_n \rangle}{\|x_n\|^2}$ converges to $\frac{1}{\lambda_k - \lambda_0}$.

13 The jacobi iteration and Seidel iterations methods aim for solving systems of linear equations iteratively.

Theorem: let $A \in M_{N\times N}$ be a regular matrix. Let $b \in \mathbb{R}^N$ be a vector and consider the system of equations

$Ax = b$

The jacobi iteration has the following iteration:

$$x_k^{(n+1)} = \frac{1}{A_{kk}}\left(-\sum_{j=1}^{k-1} A_{kj}\, x_j^{(n)} - \sum_{j=k+1}^{N} A_{kj}\, x_j^{(n)} + b_k\right) \quad k=1,2,\dots,N \quad n=0,1,2,\dots$$

formaly: decompos A into $A = L + D + U$ (lower, diagonal, and upper triangular matrices) then the jacobi iteration:

$$x^{(n+1)} = D^{-1}\left(-(L+U)\, x^{(n)} + b\right) \quad n=0,1,2,\dots$$

the jacobi iteration is convergent if the A is diagonally dominant i.e.: $|A_{kk}| > \sum_{j\neq k} |A_{kj}|$; $k=1,2,\dots,N$

for the Siedel iteration has the form:

$$x_k^{(n+1)} = \frac{1}{A_{kk}}\left(-\sum_{j=1}^{k-1} A_{kj}\, x_j^{(n+1)} - \sum_{j=k+1}^{N} A_{kj}\, x_j^{(n)} + b_k\right)$$

with $A = L + D + U \Longrightarrow x^{(n+1)} = (L+D)^{-1}\left(-Ux^{(n)} + b\right)$; $n=0,1,2,\dots$

~~the Siedel iteration~~

the Siedel iteration is Convergent if:

1) matrix A is diagonally dominant

2) matrix A is self-adjoint and positive definite