MATH280

Sammendrag av det viktige.

Markov Chains

A stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event. Here the matrix A is the probability of going from one state to another

$$\boxed{\mathbf{x}_{k+1} = A\mathbf{x}_k}$$

Example

The entries of the matrix A is the probabilities of going from one state to another (migration matrix):

$$\begin{array}{c} \text{From} \\ \begin{array}{cc} \textit{City} & \textit{Suburb} \end{array} \\ A = \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix} \begin{array}{l} \textit{City} \\ \textit{Suburb} \end{array} \end{array} \quad \text{To}$$

If we have a population vector, we can predict the comping population:

$$\boxed{A\mathbf{x}_k = \mathbf{x}_{k+1} = \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix} \begin{bmatrix} 600,000 \\ 400,000 \end{bmatrix} = \begin{bmatrix} 582,000 \\ 418,000 \end{bmatrix}}$$

PageRank is a specific case of Markov Chains.

Power method

For an $n \times n$ matrix A with an eigenvalue λ_1 bigger than all the others, then

$$\boxed{\frac{1}{\lambda_1^k} A^k \mathbf{x} \to c_1 \mathbf{v}_1}$$

where $\mathbf{v}_1$ is the eigenvector of λ_1 and c_1 is a scalar.
For a large k, a scalar multiple of $A^k\mathbf{x}$ determines the same direction as the eigenvector $c_1\mathbf{v}_1$.
This means that $A^k\mathbf{x}$ approaches the eigenspace spanned by $\mathbf{v}_1$ (as $k \to \infty$).
We do not know λ_1, therefore we have must add the constant c_1.
If we scale each $A^k\mathbf{x}$ so that its largest entry is a 1, then $\mathbf{x}_k$ converges to a multiple of $\mathbf{v}_1$ whose largest entry is 1.
When $\mathbf{x}_k$ is close to an eigenvector for λ_1 then $A\mathbf{x}_k$ is close to $\lambda_1\mathbf{x}_k$

1. Initial vector $\mathbf{x}_0$ whose largest entry is 1.
2. for $k = 0, 1, \ldots$:
   a. Compute $A\mathbf{x}_k$.
   b. Let μ_k be the largest (absolute) value in $A\mathbf{x}_k$.
   c. Compute $\mathbf{x}_{k+1} = \frac{1}{\mu_k} A\mathbf{x}_k$.
3. For almost all choises of $\mathbf{x}_0$, μ_k approaches the dominant eigenvalue and $\mathbf{x}_k$ the eigenvector.

Jacobi- and Gauss-Seidel iterative methods

We have

$$\boxed{A\mathbf{x} = \mathbf{b}}$$

where A is very large, and has a lot of zeros in it. We want to solve for $\mathbf{x}$, but doing Gaussian eliminations is expensive.
Given that it has a unique solution $\mathbf{x}^*$ and the diagonal is nonzero.

1. Initial matrix A whose diagonal entries are nonzero.
2. Write A$\mathbf{x}$ as a set of equations.
3. Solve each equation for its own x_i.
4. Guess a value for a "random" x: for instance x_1, so that the equations has a solution.
5. Loop until difference in $\mathbf{x}$ is within a tolerance:
   a. Insert the "guessed" x, and solve the set of equations.
   b. The solution will yield a new $\mathbf{x}$.
   c. Input this new $\mathbf{x}$ into the set of equations.
   d. Check the difference in the new and old $\mathbf{x}$.
   e. Repeat until the difference is below a specified tolerance.

Bases and coordinates

Let $\mathcal{B} = \{\mathbf{b}_1, \cdots, \mathbf{b}_n\}$ be a basis for V. Then each $\mathbf{x}$ in V can be expressed as a linear combination

$$\boxed{\mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_n\mathbf{b}_n}$$

The coordinates of $\mathbf{x}$ relative to $\mathcal{B}$ (or the $\mathcal{B}$-coordinates of $\mathbf{x}$), $[\mathbf{x}]_{\mathcal{B}}$, are the weights $c_1, \cdots, c_n$ such that $\mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_n\mathbf{b}_n$.

$$\boxed{[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}}$$

Change of basis

Let $\mathcal{B} = \{\mathbf{b}_1, \cdots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \cdots, \mathbf{c}_n\}$ be bases of a vector space V. Then there is a unique $n \times n$ matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ such that

$$\boxed{[\mathbf{x}]_{\mathcal{C}} = \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}[\mathbf{x}]_{\mathcal{B}}}$$

The columns of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ are the $\mathcal{C}$-coordinate vectors of the vectors in the basis $\mathcal{B}$:

$$\boxed{\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = [\ [\mathbf{b}_1]_{\mathcal{C}} \ \cdots \ [\mathbf{b}_n]_{\mathcal{C}}\ ].}$$

The matrix is the change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$. That means that multiplying by $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$

converts $\mathcal{B}$-coordinates into $\mathcal{C}$-coordinates. It also applies that $(\underset{\mathcal{C}\leftarrow\mathcal{B}}{P})^{-1} = \underset{\mathcal{B}\leftarrow\mathcal{C}}{P}$.

$$[\mathbf{c}_1\ \mathbf{c}_2 \vdots \mathbf{b}_1\ \mathbf{b}_2] \ \rightarrow\ [I \vdots \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}]$$

Linear transformations

When multiplying a vector $\mathbf{x}$ with a matrix A, we get a new vector $\mathbf{b}$.

$$A\mathbf{x} = \mathbf{b} \qquad \Leftrightarrow \qquad x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$$

This is equivalent to saying that multiplication by A transforms $\mathbf{x}$ into $\mathbf{b}$.
For each $\mathbf{x}$ in $\mathbb{R}^n$, T($\mathbf{x}$) is computed as A$\mathbf{x}$ where A $\in \mathbb{R}^{m\times n}$. This can be denoted by $\mathbf{x} \mapsto A\mathbf{x}$ ($\mathbf{x}$ maps to A$\mathbf{x}$).
The range of T is the same as Col(A).
A transformation (or mapping) T is *linear* if:

1. T($\mathbf{u}$+$\mathbf{v}$) = T($\mathbf{u}$) + T($\mathbf{v}$) for all $\mathbf{u}$, $\mathbf{v}$ in the domain of T (Col(A)).
2. T($c\mathbf{u}$) = cT($\mathbf{u}$) for all scalars c and all $\mathbf{u}$ in the domain of T (Col(A)).

Projections

Normality and orthogonality

The norm of a vector is

$$||\mathbf{v}|| = \sqrt{\mathbf{v}\cdot\mathbf{v}} \qquad \& \qquad ||\mathbf{v}||^2 = \mathbf{v}\cdot\mathbf{v}$$

The distance between two vectors are

$$\text{dist}(\mathbf{v},\mathbf{u}) = ||\mathbf{v}-\mathbf{u}|| = \sqrt{(v_1-u_1)^2 + \cdots + (v_n-u_n)^2}$$

The orthogonal compliment is the set of all vectors orthogonal to a subspace.

$$\text{Row(A)}^{\perp} = \text{Nul(A)} \qquad \& \qquad \text{Col(A)}^{\perp} = \text{Nul(A}^T\text{)}$$

Proof

The definition of the nullspace Nul(A) is the set of vectors $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{0}$. This is saying that the dot-product of these x's and rows are all zero, and is therefore orthogonal.

Orthogonal projections

Let $\mathbf{y}$ be a vector you want to project down to a subspace $\mathcal{W}$ with orthogonal basis $\mathcal{B} = \{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$, then

$$\boxed{\hat{\mathbf{y}} = \text{Proj}_{\mathcal{W}}\mathbf{y} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y}\cdot\mathbf{u}_p}{\mathbf{u}_p\cdot\mathbf{u}_p}\mathbf{u}_p}$$

Orthonormal bases

If $\mathbf{U} = \{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ form an orthonormal basis, then

$$\boxed{\hat{\mathbf{y}} = \text{Proj}_{\mathcal{W}}\mathbf{y} = \mathbf{U}\mathbf{U}^T\mathbf{y}}$$

Gram-Schmidt process

Given a basis $\{\mathbf{x}_1, \cdots, \mathbf{x}_p\}$ for a nonzero subspace $\mathcal{W}$ of $\mathbb{R}^n$, then one can form an orthogonal basis for the same subspace through these steps:

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x}_p\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{x}_p\cdot\mathbf{v}_{p-1}}{\mathbf{v}_{p-1}\cdot\mathbf{v}_{p-1}}\mathbf{v}_{p-1}
\end{aligned}$$

By doing this one creates an orthogonal basis $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ for $\mathcal{W}$.

QR-factorisation

If A $\in \mathbb{R}^{m\times n}$ with linearly independent columns, then A can be factored as $A = QR$ where Q $\in \mathbb{R}^{m\times n}$ whose columns form an orthonormal basis for Col(A) and R $\in \mathbb{R}^{n\times n}$ upper triangular invertible matrix with positive diagonal entries.
The columns of Q can be found through the Gram-Schmidt process, so that

$$\boxed{\text{Span}\{\mathbf{x}_1, \cdots, \mathbf{x}_k\} = \text{Span}\{\mathbf{u}_1, \cdots, \mathbf{u}_k\} \qquad \Rightarrow \qquad Q = [\mathbf{u}_1 \cdots \mathbf{u}_n]}$$

This means that $\mathbf{x}_k$ can be written as a linear combinations of $\mathbf{U}$:

$$\boxed{\mathbf{x}_k = r_{1,k}\mathbf{u}_1 + \cdots + r_{k,k}\mathbf{u}_k + 0\mathbf{u}_{k+1} + \cdots + 0\mathbf{u}_n}$$

This shows that $\mathbf{x}_k$ is a linear combination of the columns of Q using as weights the entries in the vector

$$\boxed{\mathbf{r}_k = \begin{bmatrix} r_{1,k} \\ \vdots \\ r_{k,k} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \& \qquad \mathbf{x}_k = Q\mathbf{r}_k \qquad \& \qquad R = [\mathbf{r}_1 \cdots \mathbf{r}_n] = Q^T A}$$

In order to find R one can use the fact that Q form an orthonormal set:

$$\boxed{R = Q^T A \qquad \Leftarrow \qquad Q^T A = Q^T(QR) = IR = R \qquad \Leftarrow \qquad Q^T Q = I}$$

Least-squares

If A $\in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$, a least squares solution of $A\mathbf{x} = \mathbf{b}$ is an $\hat{\mathbf{x}} \in \mathbb{R}^n$ which minimizes $||\mathbf{b} - A\hat{\mathbf{x}}||$.
The solution to this is to approximate $\mathbf{b}$ onto Col(A)

$$\boxed{\hat{\mathbf{b}} = \text{proj}_{\text{Col(A)}}\mathbf{b}}$$

Which makes us able to solve for the best approximation of $\mathbf{x}$

$$\boxed{A\hat{\mathbf{x}} = \hat{\mathbf{b}}}$$

Proof

Since $\mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to Col(A), then $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to each column of A. If $\mathbf{a}_j$ is a column of A, then $\mathbf{a}_j \cdot (\mathbf{b} - A\hat{\mathbf{x}}) = 0$.

$$\begin{aligned} A^T(\mathbf{b} - A\hat{\mathbf{x}}) &= \mathbf{0} \\ A^T\mathbf{b} - A^TA\hat{\mathbf{x}} &= \mathbf{0} \\ A^TA\hat{\mathbf{x}} &= A^T\mathbf{b} \end{aligned}$$

Which show that the set of least-squares solutions are found through the normal equations

$$\boxed{A^TA\mathbf{x} = A^T\mathbf{b}}$$

Spectral theorem

Properties of a $n \times n$ symmetric matrix A:

1. A has n real eigenvalues, counting duplicates.
2. The dimension of the eigenspace for eigenvalue λ is equal to the number of eigenvalues that has the same value.
3. The eigenspace for eigenvalues with different eigenvalues are orthogonal. **NOTE:** This does not apply for eigenvectors with the same eigenvalues. We can find a orthogonal space using the Gram-Schmidt method.
4. A is orthogonally diagonalizable.

Given a symmetric matrix A, we know it is orthogonally diagonalizable. And because eigenvectors will be orthogonal (if all eigenvalues are distinct) we can normalize (ending up with orthonormality). This gives us $\mathbf{A} = \mathbf{PDP}^{-1}$, where the columns of P are the normalized eigenvectors. This results in

$\mathbf{P}^{-1} = \mathbf{P}^T$, which gives:

$$\mathbf{A} = \mathbf{PDP}^{-1} = \underbrace{\begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{u}_1 & \cdots & \mathbf{u}_n \\ \downarrow & & \downarrow \end{bmatrix}}_{\text{Orthonormal eigenvectors}} \underbrace{\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}}_{\text{Eigenvalues}} \underbrace{\begin{bmatrix} \leftarrow & \mathbf{u}_1^T & \rightarrow \\ & \vdots & \\ \leftarrow & \mathbf{u}_n^T & \rightarrow \end{bmatrix}}_{\text{Transposed orthonormal eigenvectors}} = \begin{bmatrix} \uparrow & & \uparrow \\ \lambda_1\mathbf{u}_1 & \cdots & \lambda_n\mathbf{u}_n \\ \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} \leftarrow & \mathbf{u}_1^T & \rightarrow \\ & \vdots & \\ \leftarrow & \mathbf{u}_n^T & \rightarrow \end{bmatrix}$$

This can be written as:

$$\mathbf{A} = \lambda_1\mathbf{u}_1\mathbf{u}_1^T + \cdots + \lambda_n\mathbf{u}_n\mathbf{u}_n^T$$

Which is called a *spectral decomposition* of A because it breaks up A into pieces determined by the spectrum (eigenvalues) of A. This means that each matrix $\mathbf{u}_i\mathbf{u}_i^T$ is the projection matrix of a $\mathbf{x}$ onto the subspace spanned by $\mathbf{u}_i$.

Singular Value Decomposition (SVD)

If we have a matrix $\mathrm{A} \in \mathbb{R}^{m\times n}$ with rank r. There is a matrix $\Sigma \in \mathbb{R}^{m\times n}$ where the first r diagonal entries are the singular values ($\sigma_1 \geq \cdots \geq \sigma_r > 0$) of A.

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \sigma_1 & & 0 & 0 \\ & \ddots & & 0 \\ 0 & & \sigma_r & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Σ has zeros filled in for $m-r$ rows and $n-r$ columns.
There also exist two orthogonal matrices in $\mathbb{R}^{m\times m}$ (U) and $\mathbb{R}^{\times n}$ (V) such that

$$A = U\Sigma V^T = \underbrace{\begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{u}_1 & \cdots & \mathbf{u}_r \\ \downarrow & & \downarrow \end{bmatrix}}_{\text{Eigenvectors of } \mathrm{AA}^T} \underbrace{\begin{bmatrix} \sigma_1 & & 0 & 0 \\ & \ddots & & 0 \\ 0 & & \sigma_r & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{\text{Singular values } (\sqrt{\lambda_n}, \text{ with } \lambda_n \text{ eigenvalues of } \mathrm{AA}^T \text{ and } \mathrm{A}^T\mathrm{A})} \underbrace{\begin{bmatrix} \leftarrow & \mathbf{v}_1^T & \rightarrow \\ & \vdots & \\ \leftarrow & \mathbf{v}_r^T & \rightarrow \end{bmatrix}}_{\text{Eigenvectors of } \mathrm{A}^T\mathrm{A}}$$

$$U = \underbrace{\begin{matrix} \uparrow & & \uparrow \\ \mathbf{u}_1 & \cdots & \mathbf{u}_r \\ \downarrow & & \downarrow \end{matrix}}_{\text{Col(A)}} \underbrace{\begin{matrix} \uparrow & & \uparrow \\ \mathbf{u}_{r+1} & \cdots & \mathbf{u}_n \\ \downarrow & & \downarrow \end{matrix}}_{\text{Nul}(\mathrm{A}^T)} \qquad V = \underbrace{\begin{matrix} \uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_r \\ \downarrow & & \downarrow \end{matrix}}_{\text{Row(A)}} \underbrace{\begin{matrix} \uparrow & & \uparrow \\ \mathbf{v}_{r+1} & \cdots & \mathbf{v}_n \\ \downarrow & & \downarrow \end{matrix}}_{\text{Nul(A)}}$$

Weighted least-squares

Aims to minimise the sum of squared errors.

$$W\mathbf{y} = \begin{bmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} w_1y_1 \\ \vdots \\ w_ny_n \end{bmatrix}$$

We then apply the normal equation, in order to solve the least squares problem.

$$\boxed{(WA)^T WA\mathbf{x} = (WA)^T W\mathbf{y} \qquad (WA\mathbf{x} = W\mathbf{y})}$$

Minimum norm least squares

When the null space of A contains more than the zero vector, we have infinetly many solutions to a least squares problem:

$$\boxed{A^T A\mathbf{x} = A^T\mathbf{b} \quad \text{is valid for any } \mathbf{x} = \mathbf{x}^* + \mathbf{q} \text{ for any } \mathbf{q} \in Nul(A).}$$

In order to solve the normal equations for this kind of a problem, we use the pseudoinverse $A^\dagger$.

$$\boxed{A^\dagger = V_r \Sigma_r^{-1} U_r^T}$$

where r is the rank of A.

$$\boxed{\mathbf{x}^\dagger = A^\dagger\mathbf{b}}$$

Where $\mathbf{x}^\dagger$ belongs to Nul(A)$^\perp$.
(See pdf for more information and proofs.)

Tikhonov regularisation

Is a modification of the normal equations with an added regularisation parameter α.

$$\boxed{(A^T A + \alpha I)\mathbf{x}_\alpha = A^T\mathbf{b}}$$

where

$$\boxed{\mathbf{x}_\alpha = \sum_{i=1}^n \frac{\sigma_i}{\sigma_i^2+\alpha}(\mathbf{u}_i \cdot \mathbf{b})\mathbf{v}_i \qquad \lim_{\alpha\to 0} \qquad \mathbf{x}^\dagger = \sum_{i=1}^n \frac{\mathbf{u}_i\cdot\mathbf{b}}{\sigma_i}\mathbf{v}_i}$$