Mathematics for Machine Learning: Linear Algebra

Formula Sheet

Vector operations

$$\mathbf{r}+\mathbf{s}=\mathbf{s}+\mathbf{r}$$
$$2\mathbf{r}=\mathbf{r}+\mathbf{r}$$
$$\|\mathbf{r}\|^2=\sum_i r_i^2$$

- *dot or inner product*:

$$\mathbf{r}.\mathbf{s}=\sum_i r_i s_i$$

commutative $\mathbf{r}.\mathbf{s}=\mathbf{s}.\mathbf{r}$
distributive $\mathbf{r}.(\mathbf{s}+\mathbf{t})=\mathbf{r}.\mathbf{s}+\mathbf{r}.\mathbf{t}$
associative $\mathbf{r}.(a\mathbf{s})=a(\mathbf{r}.\mathbf{s})$

$$\mathbf{r}.\mathbf{r}=\|\mathbf{r}\|^2$$

$$\mathbf{r}.\mathbf{s}=\|\mathbf{r}\|\|\mathbf{s}\|\cos\theta$$

- *scalar and vector projection*:

scalar projection: $\dfrac{\mathbf{r}.\mathbf{s}}{\|\mathbf{r}\|}$

vector projection: $\dfrac{\mathbf{r}.\mathbf{s}}{\mathbf{r}.\mathbf{r}}\mathbf{r}$

Basis

A basis is a set of n vectors that:
(i) are not linear combinations of each other
(ii) span the space
The space is then n-dimensional.

Matrices

$$A\mathbf{r}=\mathbf{r}'$$
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} e \\ f \end{pmatrix}=\begin{pmatrix} ae+bf \\ ce+df \end{pmatrix}$$
$$A(n\mathbf{r})=n(A\mathbf{r})=n\mathbf{r}'$$
$$A(\mathbf{r}+\mathbf{s})=A\mathbf{r}+A\mathbf{s}$$

Identity: $I=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

clockwise rotation by θ: $\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$

determinant of 2x2 matrix: $\det\begin{pmatrix} a & b \\ c & d \end{pmatrix}=ad-bc$

inverse of 2x2 matrix: $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1}=\dfrac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

- *summation convention for multiplying matrices a and b*:

$$ab_{ik}=\sum_j a_{ij}b_{jk}$$

Change of basis

Change from an original basis to a new, primed basis. The columns of the transformation matrix B are the new basis vectors in the original coordinate system. So

$$B\mathbf{r}'=\mathbf{r}$$

where r' is the vector in the B-basis, and r is the vector in the original basis. Or;

$$\mathbf{r}'=B^{-1}\mathbf{r}$$

If a matrix A is *orthonormal* (all the columns are of unit size and orthogonal to eachother) then:

$$A^T=A^{-1}$$

Gram-Schmidt process for constructing an orthonormal basis

Start with n linearly independent basis vectors $\mathbf{v}=\{\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_n\}$. Then

$$\mathbf{e}_1=\frac{\mathbf{v}_1}{||\mathbf{v}_1||}$$

$$\mathbf{u}_2=\mathbf{v}_2-(\mathbf{v}_2.\mathbf{e}_1)\mathbf{e}_1 \quad \text{so} \quad \mathbf{e}_2=\frac{\mathbf{u}_2}{||\mathbf{u}_2||}$$

... and so on for $\mathbf{u_3}$ being the remnant part of $\mathbf{v}_3$ not composed of the preceding **e**-vectors, etc. ...

Transformation in a Plane or other object

First transform into the basis referred to the reflection plane, or whichever; E^{-1}.
Then do the reflection or other transformation, in the plane of the object T_E.
Then transform back into the original basis E.
So our transformed vector $r'=ET_EE^{-1}r$.

Eigenstuff

To investigate the characteristics of the n by n matrix $\mathbf{A}$, you are looking for solutions the the equation,

$$\mathbf{A}\mathbf{x}=\lambda\mathbf{x}$$

where λ is a scalar eigenvalue. Eigenvalues will satisfy the following condition

$$(\mathbf{A}-\lambda\mathbf{I})\mathbf{x}=0$$

where $\mathbf{I}$ is an n by n dimensional identity matrix

- *PageRank*
To find the dominant eigenvector of link matrix $\mathbf{L}$, the Power Method can be iteratively applied, starting from a uniform initial vector $\mathbf{r}$.

$$\mathbf{r}^{i+1}=\mathbf{L}\mathbf{r}^i$$

A damping factor, d, can be implement to stabilize this method as follows.

$$\mathbf{r}^{i+1}=d\mathbf{L}\mathbf{r}^i+\frac{1-d}{n}$$