Data Fundamentals (H)

Formulae

DF(H) - University of Glasgow - John H. Williamson - 2017

This lists all of the numbered formulae in Data Fundamentals (H). The most important formulae are shown with **bold titles**. You should know bolded formulae *thoroughly* .

The Unit in which the formula first appeared and is described is indicated in the section headings. The mathematical notation used is most compactly explained in Mathematical Notation: A Guide for Engineers and Scientists (https://www.amazon.co.uk/Mathematical-Notation-Guide-Engineers-Scientists/dp/1466230525).

week_1_numerical_i.ipynb

Notation for a vector

$$\mathbf{x}$$

Notation for a matrix

$$A$$

week_2_numerical_ii.ipynb

Relative precision of floating point representation.

$$\epsilon = \frac{|\text{float}(x) - x|}{|x|},$$

IEEE754 guarantee for relative precision (machine precision ϵ) for a t bit mantissa floating point number.

$$\epsilon = \frac{1}{2} 2^{-t} = 2^{-t-1},$$

week_4_matrices_i.ipynb

Addition of two vectors

$$\mathbf{x}+\mathbf{y} = [x_1+y_1, x_2+y_2, \dots, x_n+y_n]$$

Scalar multiplication a vector

$$c\mathbf{x} = [cx_1, cx_2, \dots, cx_n]$$

Linear interpolation of two vectors.

$$\text{lerp}(\mathbf{x}, \mathbf{y}, \alpha) = \alpha\mathbf{x} + (1-\alpha)\mathbf{y}$$

Cosine of angle between two vectors in terms of normalised dot product

$$\cos\theta = \frac{\mathbf{x} \bullet \mathbf{y}}{||\mathbf{x}||\ ||\mathbf{y}||}.$$

Dot product / inner product

$$\mathbf{x} \bullet \mathbf{y} = \sum_i x_i y_i.$$

Mean vector

$$\text{mean}(\mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_n}) = \frac{1}{N}\sum_i \mathbf{x_i}$$

Definition of linearity (for a linear function f and equivalent matrix A)

$$\begin{aligned} f(\mathbf{x}+\mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y}) &= A(\mathbf{x}+\mathbf{y}) = A\mathbf{x} + A\mathbf{y}, \\ f(c\mathbf{x}) = cf(\mathbf{x}) &= A(c\mathbf{x}) = cA\mathbf{x}, \end{aligned}$$

Matrix addition

$$A+B = \begin{bmatrix} a_{1,1}+b_{1,1} & a_{1,2}+b_{1,2} & \dots & a_{1,m}+b_{1,m} \\ a_{2,1}+b_{2,1} & a_{2,2}+b_{2,2} & \dots & a_{2,m}+b_{2,m} \\ \dots \\ a_{n,1}+b_{n,1} & a_{n,2}+b_{n,2} & \dots & a_{n,m}+b_{n,m} \end{bmatrix}$$

Scalar matrix multiplication

$$cA = \begin{bmatrix} ca_{1,1} & ca_{1,2} & \dots & ca_{1,m} \\ ca_{2,1} & ca_{2,2} & \dots & ca_{2,m} \\ \dots \\ ca_{n,1} & ca_{n,2} & \dots & ca_{n,m} \end{bmatrix}$$

Matrix multiplication

$$C_{ij} = \sum_k a_{ik} b_{kj}$$

Outer product (matrix version)

$$\mathbf{x} \otimes \mathbf{y} = \mathbf{x}^T\mathbf{y}$$

Inner product (matrix version)

$$\mathbf{x} \bullet \mathbf{y} = \mathbf{x}\mathbf{y}^\mathbf{T}$$

week_5_matrices_ii.ipynb

Power iteration for leading eigenvalue

$$x_n = \frac{Ax_{n-1}}{\|Ax_{n-1}\|_\infty}$$

Definition of eigenvalue and eigenvector

$$A\mathbf{x_i} = \lambda_i \mathbf{x_i}$$

Trace of a matrix

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} \lambda_i$$

Determinant of a matrix

$$\det(A) = \prod_{i=1}^{n} \lambda_i$$

Linear system Ax = y

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

SVD definition

$$A = U\Sigma V,$$

(pseudo-)inverse by SVD

$$A = U\Sigma V$$
$$A^{-1} = V^T \Sigma^\dagger U^T$$

week_6_optimisation_i.ipynb

Definition of optimisation

$$\theta^* = \underset{\theta\in\Theta}{\arg\min}\, L(\theta)$$

Objective function for approximation

$$L(\theta) = \|y' - y\| = \|f(\mathbf{x};\theta) - y\|$$

Equality constraints for optimisation

$$\theta^* = \underset{\theta\in\Theta}{\arg\min}\, L(\theta) \text{ subject to } c(\theta) = 0$$

Inequality constraints for optimisation

$$\theta^* = \underset{\theta\in\Theta}{\arg\min}\, L(\theta) \text{ subject to } c(\theta) \leq 0$$

week_7_optimisation_ii.ipynb

Gradient vector

$$\nabla L(\theta) = \left[\frac{\partial L(\theta)}{\partial \theta_1}, \frac{\partial L(\theta)}{\partial \theta_2}, \ldots, \frac{\partial L(\theta)}{\partial \theta_n}, \right]$$

∇ is called "nabla" or "del" and is used to represent the gradient of a function.

Gradient descent algorithm

$$\theta^{(\mathbf{i+1})} = \theta^{(\mathbf{i})} - \delta \nabla L(\theta^{(\mathbf{i})}),$$

Definition of differentiation

$$\frac{df(x)}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x-h)}{2h}$$

Hessian matrix

$$\nabla \nabla L(\theta) = \nabla^2 L(\theta) = \begin{bmatrix} \frac{\partial^2 L(\theta)}{\partial \theta_1^2} & \frac{\partial^2 L(\theta)}{\partial \theta_1 \partial \theta_2} & \frac{\partial^2 L(\theta)}{\partial \theta_1 \partial \theta_3} & \ldots & \frac{\partial^2 L(\theta)}{\partial \theta_1 \partial \theta_n} \\ \frac{\partial^2 L(\theta)}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 L(\theta)}{\partial \theta_2^2} & \frac{\partial^2 L(\theta)}{\partial \theta_2 \partial \theta_3} & \ldots & \frac{\partial^2 L(\theta)}{\partial \theta_2 \partial \theta_n} \\ \ldots \\ \frac{\partial^2 L(\theta)}{\partial \theta_n \partial \theta_1} & \frac{\partial^2 L(\theta)}{\partial \theta_n \partial \theta_2} & \frac{\partial^2 L(\theta)}{\partial \theta_n \partial \theta_3} & \ldots & \frac{\partial^2 L(\theta)}{\partial \theta_n^2} \end{bmatrix},$$

Convex sum of sub-objective functions

$$L(\theta) = \lambda_1 L_1(\theta) + \lambda_2 L_2(\theta) + \cdots + \lambda_n L_n(\theta),$$

Week_8_probability_I.ipynb

Boundedness of probabilities

$$0 \leq P(A) < 1$$

Unitarity of probabilities

$$\sum_A P(A) = 1$$

Sum rule for probabilities

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B),$$

$\wedge$ is "and" and $\vee$ is "or"

Conditional probability definition

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

$|$ is read as "given" or "conditioned on", as in "probability of A given B".

Empirical distribution

$$P(X = x) = \frac{n_x}{N},$$

Conditional probability in terms of joint and marginal

$$P(X|Y) = \frac{P(X,Y)}{P(Y)}$$

$$f_X(X|Y) = \frac{f_X(X,Y)}{f_X(Y)}$$

Entropy of a random variable

$$H(X) = \sum_x -P(X = x)\log_2(P(X = x))$$

Bayes Rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes Rule (in terms of hypotheses H and data D)

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

Integration over the evidence

$$P(B) = \sum_i P(B|A_i)P(A_i)$$

$$P(B) = \int_A P(B|A)P(A)dA$$

Definition of odds

$$\text{odds} = \frac{1-p}{p}$$

Definition of log-odds (logit)

$$\operatorname{logit}(p) = \log\left(\frac{p}{1-p}\right)$$

Probability of multiple independent outcomes

$$P(X_1 = x_i, \dots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i)$$

Log-probability of multiple independent outcomes

$$\log P(x_1, \dots, x_n) = \sum_{i=1}^{n} \log P(x_i)$$

Probability of continuous random variable in terms of PDF $f_X(x)$

$$P(X \in [a,b]) = P(a < X < b) = \int_a^b f_X(x)$$

Definition of cumulative density function

$$c_X(x) = \int_{-\infty}^{x} f_X(x) = P(X \leq x)$$

Notation for normally distributed variable X

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

(note $\sim$ is read "distributed as")

week_9_probability_ii.ipynb

Expectation of a random variable

$$\mathbb{E}[X] = \int_x x\ f_X(x) dx$$

Expectation of a function of a random variable

$$\mathbb{E}[g(X)] = \int_x f_X(x) g(x) dx$$

Acceptance probability for Metropolis-Hastings jump from x to x'

$$P(\text{accept}) = \begin{cases} f_X(x')/f_X(x), & f_X(x) > f_X(x') \\ 1, & f_X(x) \leq f_X(x') \end{cases}$$

week_10_signals.ipynb

Definition of Nyquist limit

$$f_n = \frac{f_s}{2}$$

Signal to noise ratio

$$\mathrm{SNR} = \frac{S}{N},$$

$$\mathrm{SNR}_{dB} = 10\log_{10}\left(\frac{S}{N}\right)$$

Exponential smoothing

$$y[t] = \alpha y[t-1] + (1-\alpha)x[t]$$

Convolution of sampled signals

$$(x*y)[n] = \sum_{m=-M}^{M} x[n]y[n-m]$$