Machine learning => math refresher

① $f(x,y,z) = x^t A y + Bx - y^t C Z D - y^t E^t y + F$

$x \in \mathbb{R}^{n}, y \in \mathbb{R}^{m}, z \in \mathbb{R}^{p \times q}, f(x,y,z) \in \mathbb{R}$

$\Rightarrow x^t A y \in \mathbb{R} = [1 \dots n] \begin{bmatrix} & \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ m \end{bmatrix} \Rightarrow A \in \mathbb{R}^{n \times m}$

$\Rightarrow Bx \in \mathbb{R} \Rightarrow B \in \mathbb{R}^{1 \times n}$

$\Rightarrow y^t C Z D \in \mathbb{R} \Rightarrow C \in \mathbb{R}^{m \times p}$

$\Rightarrow D \in \mathbb{R}^{q \times 1}$

$\Rightarrow y^t E^t y \in \mathbb{R} \Rightarrow E \in \mathbb{R}^{m \times m}$

$\Rightarrow F \in \mathbb{R}$

② $f(x) = \sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j M_{ij} = \sum_{i=1}^{N} x_i \sum_{j=1}^{N} x_j M_{ij} = \sum_{i=1}^{N} x_i (x^t M)_i = \sum_{i=1}^{N} x_i (M^t x)_i =$

required in order to return a real value

$= x^t M^t x = (x^t M^t x)^t = x^t M x$ //

③

a) for there to be a unique solution, the matrix A has to have a rank equal to the dimension equal to the number of variables in the x vector. Implying that there are enough linearly independent columns/rows in order to solve the equations for every variable

b) given that there are 5 variables to solve, the matrix should have rank 5 for there to be a unique solution. On the other hand, the rank is equal to the number of non-zero eigenvalues of the matrix A (4). therefore, there are no unique solutions to this system of linear equations.

④ $A, B \in \mathbb{R}^{n \times n}$

$AB = BA = I$, A is inverse to B $\Rightarrow$ $\det A \neq 0$

$\Rightarrow$ $\lambda_i \neq 0 \in i$ in eigenvalues of A

⑤ $x^t A x \geq 0$ $\Rightarrow$ PSD

no negative eigenvalues $\Leftrightarrow$ PSD

$\Rightarrow \lambda_i \neq 0$ for $i \in$ eigenvalues A

since A is symmetrical, it's eigenvectors will be orthogonal with each other (U^{-1} exists)

, $\Rightarrow A = U \Lambda U^{-1}$ where U is the eigenvector matrix

and $\Lambda = \mathrm{diag}(\lambda_i)$ for $\lambda_i \in A$'s eigenvalues

$= U \Lambda U^t$ (since an inverted orthogonal matrix is equal to the original matrix transposed)

which means that $x^t A x = x^t U \Lambda U^t x = y^t \Lambda y = \sum_{i=1}^{n} \lambda_i y_i^2 \geq 0$ if $\lambda_i \geq 0$

$\hookrightarrow y = U^t x$

(possible since U is full rank)

⑥ $A \in \mathbb{R}^{m \times n}$

prove: $B = A^t A$ is PSD

$\Leftrightarrow$ $x^t B x \geq 0$ for all x

, $x^t B x = x^t A^t A x = \overset{\text{vector}}{\widetilde{(Ax)^t}} A x$

$= (Ax)^t \, Ax$ —o inner product

$= \| Ax \|^2 \geq 0$ //

⑦ $f(x) = 1/2 \, a x^2 + b x + c$

$\min_{x \in \mathbb{R}} f(x)$?

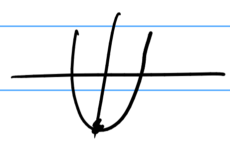
a) 1 solution: $a > 0$

infinitely: $a \leq 0$

zero : $a = b = 0$, $c \in \mathbb{R}$

b) since there is a unique solution $\Rightarrow$ a is positive $\Rightarrow$ the parabola is facing upwards

the minimum occurs at $\dfrac{-d}{2e}$ where $d = b$ and $e = c$



$\Rightarrow \arg\min_{x \in \mathbb{R}} f(x) = \dfrac{-b}{2c}$

8 ⓑ $g(x) = \frac{1}{2} x^t A x + b^t x + c$ where A is symmetric and PSD

$\nabla b^t x = b$

$$\nabla^2 g(x) = \nabla^2 \tfrac{1}{2} x^t A x + \nabla^2 b^t x + \nabla^2 c$$
$$= \nabla^2 \tfrac{1}{2} x^t A x + \nabla^2 b^t x + 0 \qquad \nabla_x(\nabla_x b^t x) = \nabla_x \left(\frac{\partial}{\partial x_k} \sum_{i=1}^{m} b_i x_i \right) = \nabla_x(b) = 0$$

- $\nabla_x(\nabla_x b^t x)$, each entry of $\nabla_x(b^t x)$ is $\frac{\partial}{\partial x} \sum_{i=1}^{m} b_i x_i$

$$\Leftrightarrow \nabla_x(b^t x) = b$$
$$\Rightarrow \nabla_x(b) = \nabla_x(\nabla_x b^t x) = 0$$

- $\nabla_x(\nabla_x \frac{1}{2} x^t A x)$, each entry of $\nabla_x(x^t A x)$ is $\frac{\partial}{\partial x_k} \sum_{i=1}^{m} \sum_{j=1}^{m} A_{ij} x_i x_j$

$$= \frac{\partial}{\partial x_k} \left(\sum_{i \neq k} \sum_{j \neq k} A_{ij} x_i x_j + \sum_{i \neq k} A_{ik} x_i x_k + \sum_{j \neq k} A_{kj} x_k x_j + A_{kk} x_k^2 \right)$$

$$= 0 + \sum_{i \neq k} A_{ik} x_i + \sum_{j \neq k} A_{kj} x_j + 2 A_{kk} x_k = \sum_{i=1}^{m} A_{ik} x_i + \sum_{j=1}^{m} A_{kj} x_j$$

A is symmetric

$$= 2 \sum_{i=1}^{m} A_{ki} x_i \qquad \Rightarrow \nabla_x(x^t A x) = 2Ax$$

each entry of $\nabla_x(Ax)$ is $\frac{\partial^2 Ax}{\partial x_k x_l} = \frac{\partial}{\partial x_k} \left(2 \sum_{i=1}^{m} A_{li} x_i \right) = 2 A_{lk} = 2 A_{kl}$

$$\Rightarrow \nabla_x(\nabla_x(x^t A x)) = 2A \;//$$

⑨ $p(a|b,c) = p(a|c) \Rightarrow p(a|b) = p(a)$

say that b is a subset of c, then we have that $p(a|b,c) = p(a|c)$, however this doesn't imply that $p(a|b) = p(a)$ because the probability of the event a may as well change by knowing b beforehand.

(10) $p(a|b) = p(a) \Rightarrow p(a|b,c) = p(a|c)$

$$\bullet\ p(a|b,c) = \frac{p(a,b,c)}{p(b,c)} = \frac{p(b|a,c)\,p(a,c)}{p(b,c)} = \frac{p(b|a,c)\,p(a|c)\,\cancel{p(c)}}{p(b|c)\,\cancel{p(c)}}$$

$$= \frac{p(b|a,c)\,p(a|c)}{p(b|c)} = p(a,c)$$

$\hookrightarrow p(b|a,c) = p(b|c)$
(a and b are independent)

(11) $\cdot\ p(c)$, $\quad p(a,b,c) = p(a|b,c)\,p(b,c) = \frac{p(b,c|a)\,p(a)}{p(\cancel{b,c})}\,p(\cancel{b},c) =$

$$= p(b,c|a)\,p(c) \Leftrightarrow p(c) = \frac{p(a,b,c)}{p(b,c|a)}$$

$\cdot\ p(c|a,b)$, $\quad p(a,b,c) = p(a,b|c)\cdot p(c) = \frac{p(c|a,b)\,p(a,b)\,\cancel{p(c)}}{p(\cancel{c})}$

$$= p(c|a,b)\,p(a,b) \Leftrightarrow p(c|a,b) = \frac{p(a,b,c)}{p(a,b)}$$

$\cdot\ p(b|c)$, $\quad p(a,b,c) = p(a|b,c)\cdot p(b,c) = p(a|b,c)\,p(b|c)\,p(c)$

$$\Leftrightarrow p(b|c) = \frac{p(a,b,c)}{p(a|b,c)\,p(c)}$$

(12) $p(p|s) = 0{,}95$

$p(\bar{p}|\bar{s}) = 0{,}95$

$p(s) = 0{,}001$

$1/1000$

bayes theorem

$$p(s|p) = \frac{p(p|s)\,p(s)}{p(p)} = \frac{0{,}95 \times 0{,}001}{0{,}0505} \simeq 0{,}01866$$

$$p(p) = p(p|s)\,p(s) + p(p|\bar{s})\,p(\bar{s})$$

$$= 0{,}95 \times 0{,}001 + (1 - p(\bar{p}|\bar{s}))\,(1 - p(s))$$

$$= 0{,}00095 + 0{,}05 \times 0{,}999 = 0{,}0505$$

(13) $X \sim N(\mu, \sigma^2) \Rightarrow pdf(x) = \frac{1}{2\pi\sigma} \exp(-1/2\sigma^2 \cdot (x-\mu)^2)$

$f(x) = ax + bx^2 + c$

$$E(f(x)) = \int_{-\infty}^{\infty} pdf(x) f(x)\, dx = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma} \exp(-1/2\sigma^2 \cdot (x-\mu)^2)(ax + bx^2 + c)\, dx$$

(14) $$E[g(x)] = \int_{\mathbb{R}^n} g(x) f_{x_1, \dots, x_m}(x_1, \dots, x_n; \mu, \Sigma) =$$

$$= \int_{\mathbb{R}^n} Ax \frac{1}{2\pi^{n/2} |\Sigma|^{1/2}} e^{\left(-1/2(-x-\mu)^T \Sigma^{-1} (x-\mu)\right)} dx_1 \dots dx_m$$

$$E[g(x) g(x)^T] = \int_{\mathbb{R}^n} A\|x\|^2 A^T \frac{1}{2\pi^{n/2} |\Sigma|^{1/2}} e^{\left(-1/2(-x-\mu)^T \Sigma^{-1} (x-\mu)\right)} dx_1 \dots dx_m$$

$$E[g(x)^T g(x)] = \int x^T A^T A x \cdot \frac{1}{2\pi \Sigma_{11}} e^{-1/2\Sigma_{11} \cdot (x-\mu_1)^2} dx$$

$\rightarrow Cov[g(x)]$