1 Probability and Statistics

(1) (combinatorics)

Evidently, C(1, 0) = C(1, 1) = 1

Suppose for N = n ∈ N^+, that for all 0 <= k <= n, we have

C(n, k) = n! / (k! * (n - k)!)

Then let N = n + 1, for all 0 <= k+1 <= n, from the property of C, we have

$$C(n+1, k+1) = C(n, k+1) + C(n, k) = \frac{n!}{(k+1)!\,(n-k-1)!} + \frac{n!}{k!\,(n-k)!}$$

$$= \frac{n!\,((n-k)+(k+1))}{(k+1)!\,(n-k)!} = \frac{(n+1)!}{(k+1)!\,(n-k)!}$$

And for k = 0, C(n, 0) = 1 still holds.

Thus for N = n + 1, the equation still holds.

Therefore, for any N∈ N^+ and any 0 <= K <= N, we have

$$C(N, K) = \frac{N!}{K!\,(N-K)!}$$

(2) (counting)

(a) Altogether 2^{10} cases, each with probability $(1/2)^{10}$.

Among them, C(10, 4) cases fit our demand.

P(head = 4, tail = 6) = C(10, 4) * $(1/2)^{10}$

(b) Altogether 2^{10} cases, each with probability 1 / C(52, 5).

Among them, take 2 numbers, from each number take 3/2 cards.

X – Y are commutative, thus the result is multiplied by 2.

P(full house) = $2 * \frac{C(13,2)C(4,2)C(4,3)}{C(52,5)}$

(3) (conditional probability)

$$P(\text{head} = 3 \mid \text{head} \geq 1) = \frac{P(\text{head} \geq 1 \mid \text{head} = 3)P(\text{head} = 3)}{P(\text{head} \geq 1)} = \frac{1 * \left(\frac{1}{8}\right)}{\left(1 - \left(\frac{1}{8}\right)\right)} = \frac{1}{7}$$

(4) (Bayes theorem)

$$P(X = -1 \mid |X| = 1) = \frac{P(|X| = 1 \mid X = -1)\,P(X = -1)}{P(|X| = 1)} = \frac{1 * \left(\frac{1}{2}\right) * \left(\frac{1}{4}\right)}{\left(\frac{1}{2}\right) * \left(\frac{1}{4}\right) + \left(\frac{1}{2}\right) * \left(\frac{1}{8}\right)} = \frac{2}{3}$$

(5) (union/intersection)

(a) Max(P(A ∩ B)) = 0.3 when A ⊆ B

(b) Min(P(A ∩ B)) = 0 when A ∩ B = ∅

(c) Max(P(A ∪ B)) = 0.7 when A ∩ B = ∅

(d) Min(P(A ∪ B)) = 0.4 when A ⊆ B

2 Linear Algebra

(1) (rank)

It is a rank-2 square matrix.

(2) (inverse)

$$\begin{matrix} 0.125 & -0.625 & 0.75 \\ -0.25 & 0.75 & -0.5 \\ 0.375 & -0.375 & 0.25 \end{matrix}$$

(3) (eigenvalues/eigenvectors)

eigenvalues	eigenvectors
4	1, 2, -1
2	1, 0, -1
2	0, 1, -1

(4) (singular value decomposition)

(a) $MM^{\dagger}M = U\Sigma V^T V\Sigma^{\dagger}U^T U\Sigma V^T = U\Sigma\Sigma^{\dagger}\Sigma V^T$

$M = U\Sigma V^T$

Obviously, $MM^{\dagger}M = M$ if $\Sigma\Sigma^{\dagger}\Sigma = \Sigma$

w.l.o.g Suppose M is an m-by-n (m<n) matrix, then Σ is an m-by-n matrix with non-negative real numbers on the diagonal.

$$\Sigma = \begin{pmatrix} \sigma_1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \sigma_m & 0 \end{pmatrix}$$

$$\Sigma^{\dagger} = \begin{pmatrix} \sigma_1^{\dagger} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_m^{\dagger} \\ 0 & \ldots & 0 \end{pmatrix}, \text{ where } \sigma_i^{\dagger} = \begin{cases} 1/\sigma_i, & if\ \sigma_i \neq 0 \\ 0, & otherwise \end{cases}, 1 \leq \text{i} \leq \text{m}$$

Thus,

$$\Sigma\Sigma^{\dagger} = \begin{pmatrix} \sigma_1^* & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_m^* \end{pmatrix}, \text{ where } \sigma_i^* = \begin{cases} 1, & if\ \sigma_i \neq 0 \\ 0, & otherwise \end{cases}, 1 \leq \text{i} \leq \text{m}$$

$$\Sigma\Sigma^{\dagger}\Sigma = \begin{pmatrix} \sigma_1\sigma_1^* & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \sigma_m\sigma_m^* & 0 \end{pmatrix}, \text{ where } \sigma_i\sigma_i^* = \begin{cases} \sigma_i, & if\ \sigma_i \neq 0 \\ 0, & otherwise \end{cases}, 1 \leq \text{i} \leq \text{m}$$

Therefore for any $1 \leq \text{i} \leq \text{m}, 1 \leq \text{j} \leq \text{n}$, we have

$(\Sigma\Sigma^{\dagger}\Sigma)[i][j] = \Sigma[i][j]$, which means $\Sigma\Sigma^{\dagger}\Sigma = \Sigma$

(b) w.l.o.g Suppose M is an m-by-n (m<n) matrix, then Σ is an m-by-n matrix with non-negative real numbers on the diagonal.

M is invertible, therefore rank(M) = m. Since U_{mxm} and V_{nxn} are orthogonal, rank(U) = m, rank(V) = n > m, thus rank(Σ) >= m, which suggests $\sigma_1, \ldots, \sigma_m \neq 0$

Similar to the proof in (a), we have

$$\Sigma\Sigma^{\dagger} = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix} = I$$

Therefore $MM^{\dagger} = U\Sigma V^T V \Sigma^{\dagger} U^T = I$

This indicates that $M^{\dagger} = M^{-1}$

(5) (PD/PSD)

(a) for all x ≠ 0, $xZZ^T x^T = xZ(xZ)^T \geq 0$, therefore ZZ^T is PSD

(b) =>

Let λ, v be an eigenvalue-eigenvector pair of A.

Then $Av = \lambda v$

Given that A is PD, there exists at least one $v \neq \mathbf{0}$, and $v^T Av = \lambda v^T v > 0$

Since $v^T v > 0$, λ has to be strictly positive.

<=

Suppose there exists an eigenvalue-eigenvector pair λ, v of A s.t. λ <= 0

Then $v^T Av = \lambda v^T v \leq 0$, which contradicts with the fact that A is PD.

This indicates that each eigenvalue of A has to be strictly positive.

(6) (inner product)

(a) $\max(\boldsymbol{u}^T \boldsymbol{x}) = ||\boldsymbol{x}||, when\ \boldsymbol{u} = \frac{\boldsymbol{x}}{||\boldsymbol{x}||}$

(b) $\min(\boldsymbol{u}^T \boldsymbol{x}) = -||\boldsymbol{x}||, when\ \boldsymbol{u} = -\frac{\boldsymbol{x}}{||\boldsymbol{x}||}$

(c) $\min(|\boldsymbol{u}^T \boldsymbol{x}|) = 0$. If $\boldsymbol{x} = \mathbf{0}$ apparently any $\boldsymbol{u}$ is good. Otherwise,

let $\boldsymbol{x} = (a_1, \dots, a_d)$, w.l.o.g suppose $a_1 \neq 0$

let $\boldsymbol{v} = (b_1, \dots, b_d), where\ \sum_{i=2}^{d} b_i^2 \neq 0, b_1 = \frac{\sum_{i=2}^{d} a_i b_i}{a_1}$

Evidently, $\boldsymbol{v}^T \boldsymbol{x} = 0$

So we just make $\boldsymbol{u} = \frac{\boldsymbol{v}}{||\boldsymbol{v}||}$

3 Calculus

(1) (differential and partial differential)

$$\frac{df(x)}{dx} = \frac{-2e^{-2x}}{1+e^{-2x}}$$

$$\frac{\partial g(x,y)}{\partial y} = 2e^{2y} + 6xye^{3xy^2}$$

(2) (chain rule)

$$\frac{\partial f}{\partial v} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v} = -y\sin(u+v) - x\cos(u-v)$$

(3) (gradient and Hessian)

$$\nabla E = \boldsymbol{i}\frac{\partial E}{\partial u} + \boldsymbol{j}\frac{\partial E}{\partial v}, where$$

$$\frac{\partial E}{\partial u} = 2(ue^v - 2ve^{-u})(e^v + 2ve^{-u})$$

$$\frac{\partial E}{\partial v} = 2(ue^v - 2ve^{-u})(ue^v - 2e^{-u})$$

At u = 1 and v = 1, $\nabla E = 2(e^2 - 4e^{-2})\boldsymbol{i} + 2(e - 2e^{-1})^2\boldsymbol{j}$

$$\nabla^2 E = \begin{pmatrix} \frac{\partial^2 E}{\partial^2 u} & \frac{\partial^2 E}{\partial u \partial v} \\ \frac{\partial^2 E}{\partial v \partial u} & \frac{\partial^2 E}{\partial^2 v} \end{pmatrix}, where$$

$$\frac{\partial^2 E}{\partial^2 u} = 2(ue^v - 2ve^{-u})^2 + 2(ue^v - 2ve^{-u})(-2ve^{-u})$$

$$\frac{\partial^2 E}{\partial u \partial v} = 2(e^v + 2ve^{-u})(ue^v - 2e^{-u}) + 2(ue^v - 2ve^{-u})(e^v + 2e^{-u})$$

$$\frac{\partial^2 E}{\partial v \partial u} = 2(e^v + 2ve^{-u})(ue^v - 2e^{-u}) + 2(ue^v - 2ve^{-u})(e^v + 2e^{-u})$$

$$\frac{\partial^2 E}{\partial^2 v} = 2(ue^v - 2e^{-u})^2 + 2(ue^v - 2ve^{-u})(ue^v)$$

At u = 1 and v = 1, $\nabla^2 E = \begin{pmatrix} 2e^2 - 12 + 16e^{-2} & 4e^2 - 16e^{-2} \\ 4e^2 - 16e^{-2} & 4e^2 - 4 + 8e^{-2} \end{pmatrix}$

(4) (Taylor's expansion)

$E(1 + \Delta u, 1 + \Delta v)$

$$= E(1,1) + \Delta u \frac{\partial E(1,1)}{\partial u} + \Delta v \frac{\partial E(1,1)}{\partial v}$$
$$+ \frac{1}{2!}\left[(\Delta u)^2 \frac{\partial^2 E(1,1)}{\partial^2 u} + 2\Delta u \Delta v \frac{\partial^2 E(1,1)}{\partial u \partial v} + (\Delta v)^2 \frac{\partial^2 E(1,1)}{\partial^2 v}\right] + o^2$$
$$= (e - 2e^{-1})^2 + \Delta u * 2(e^2 - 4e^{-2}) + \Delta v * 2(e - 2e^{-1})^2$$
$$+ \frac{1}{2!}[(\Delta u)^2 * 2e^2 - 12 + 16e^{-2} + 2\Delta u \Delta v * (4e^2 - 16e^{-2})$$
$$+ (\Delta v)^2(4e^2 - 4 + 8e^{-2})] + o^2$$

(5) (optimization)

$F(\alpha) = Ae^{\alpha} + Be^{-2\alpha}$

Let $\frac{dF}{d\alpha} = Ae^{\alpha} - 2Be^{-2\alpha} = 0$, since A > 0, B > 0, we have $\alpha^* = \frac{\ln(2B) - \ln A}{3}$

$\frac{dF(\alpha^{*+})}{d\alpha} > 0, \frac{dF(\alpha^{*-})}{d\alpha} < 0$, therefore $F(\alpha^*)$ is the minimum.

(6) (vector calculus)

Let $\boldsymbol{w} = (w_1, \dots, w_d)$, $\boldsymbol{b} = (b_1, \dots, b_d)$, $\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1d} \\ \vdots & \ddots & \vdots \\ a_{d1} & \cdots & a_{dd} \end{pmatrix}$

$\boldsymbol{w}^T \mathbf{A} \boldsymbol{w} = w_1^2 a_{11} + \cdots + w_1 w_d a_{1d} + \cdots + \cdots + w_d w_1 a_{d1} + \cdots + w_d^2 a_{dd}$

$\boldsymbol{b}^T \mathbf{w} = b_1 w_1 + \cdots + b_d w_d$

Given that **A** is symmetrical, for 1<=i<=d

$$\frac{\partial \boldsymbol{w}^T \mathbf{A} \boldsymbol{w}}{\partial w_i} = 2 \sum_{k=1}^{d} a_{ik} w_i$$

Therefore

$$\frac{\partial E}{\partial w_i} = \sum_{k=1}^{d} a_{ik} w_i + b_i$$

Which yields

$$\frac{\partial^2 E}{\partial w_i \partial w_j} = a_{ij}$$

Henceforth

$$\nabla E = \frac{\partial E}{\partial \boldsymbol{w}} = \mathbf{A}\boldsymbol{w} + \boldsymbol{b}$$

$$\nabla^2 E = \frac{\partial^2 E}{\partial^2 \boldsymbol{w}} = \mathbf{A}$$