S

$0, 1, \cdots, 9$

$$S = \{0, 1, \cdots, 9\}$$

$A \subset S$

$\emptyset$

$$A = \{0,\ 2,\ 4,\ 6,\ 8\} \subset S = \{0,\ 1, \cdots, 9\}$$

A

$$s \in A$$

$P(A)$

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

$$P(\emptyset) = 0$$

$$P(S) = 1$$

$$A = \{0, 1, 2, 3\} \subset S$$

$$P(A) = 4/10 = 0.4$$

$(S,\, A,\, P)$

P

$$x = x(s)$$

$s \in S$

x

$$F_X(u_1, \cdots, u_N)$$

=

$$P(x_1 < u_1, \cdots, x_N < u_N)$$

$$\int_{-\infty}^{u_1} \cdots \int_{-\infty}^{u_N} p(\xi_1, \cdots, \xi_N) d\xi_1 \cdots d\xi_N$$

$$p(\xi_1, \cdots, \xi_N)$$

N

y

$$P(A \cap B) = P(x < u, y < v) = P(x < u)P(y < v) = P(A)P(B)$$

B

$$x < u$$

$y < v$

$$p(x, y) = p(x)p(y)$$

$$P(x < u, y < v)$$

$$\int_{-\infty}^{u}\int_{-\infty}^{v} p(\xi,\eta)d\xi d\eta = \int_{-\infty}^{u}\int_{-\infty}^{v} p(\xi)p(\eta)d\xi d\eta$$

$$\int_{-\infty}^{u} p(\xi)d\xi \int_{-\infty}^{v} p(\eta)d\eta = P(x < u)P(y < v)$$

$$p(x_1, \cdots, x_N) = p(x_1)\, p(x_2) \cdots p(x_N)$$

$$P(A|B)$$

$$P(A|B) = P(A)$$

$$A \cup B$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$A \cap B = \emptyset$$

$$P(A \cup B) = P(A) + P(B)$$

$$A \cap B$$

$$P(A, B) = P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$$

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

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

$$\left.\begin{array}{l} P(A|B)\, P(B|C) = P(B|A)\, P(A|C) \\ P(A|B,C)\, P(B|C) = P(B|A,C)\, P(A|C) \end{array}\right\} = P(A,B|C)$$

$$P(A|B) = P(A),$$

$$P(B|A) = P(B)$$

$$P(A, B) = P(A)P(B)$$

$$P(A=a) = \int P(A=a,\, B=b)\, db = \int P(a|b)P(b)\, db$$

$\{A_1, \cdots, A_n\}$

n

$$\bigcup_{i=1}^{n} A_i = S,$$

$$A_i \cap A_j = \emptyset \quad (i \neq j)$$

$B \subset S$

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i)P(A_i)}{P(B)}$$

$$P(B) = P(B \cap S) = P(B \cap (\bigcup_{i=1}^{n} A_i)) = P(\bigcup_{i=1}^{n}(B \cap A_i)) = \sum_{i=1}^{n} P(B \cap A_i)$$

$$P(B)$$

$$F_x(\xi) = P(x < \xi)$$

$$P(a \leq x < b) = F_x(b) - F_x(a)$$

$p(x)$

$$F_x(\xi) = \int_{-\infty}^{\xi} p(x)dx$$

$$p(x) = \frac{d}{d\xi} F_x(\xi)$$

$$P(a \leq x < b) = F_x(b) - F_x(a) = \int_a^b p(x)dx$$

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

$$\{x_i \quad i = 1, \cdots, n\}$$

$$P(x = x_i) = P(x_i) = P_i \quad (i = 1, \cdots, n)$$

$$0 \leq P_i \leq 1, \qquad \sum_{i=1}^{n} P_i = 1$$

$$F_x(\xi) = P(x < \xi) = \sum_{x_i < \xi} P(x_i) = \sum_{i=1}^{k} P_i$$

$$x_1 < \cdots < x_n$$

$$\xi = x_{k+1}$$

$$\mu_x = E(x) = \sum_{i=1}^{n} x_i P(x_i) = \sum_{i=1}^{n} x_i P_i$$

$$\mu_x = E(x) = \int_{-\infty}^{\infty} x\ p(x)dx$$

$$\{x_1, \cdots, x_n\}$$

$$E(x)$$

$$\hat{\mu}_x = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$\tilde{x}$

$$\tilde{x} = \begin{cases} x_{(n+1)/2} & \text{if } n \text{ is odd} \\ \frac{x_{n/2}+x_{n/2+1}}{2} & \text{if } n \text{ is odd} \end{cases}$$

$$\int_{-\infty}^{\tilde{x}} p(x)dx = \int_{\tilde{x}}^{\infty} p(x)dx = 0.5$$

$$\tilde{x} = \arg\max_{x} p(x),$$

$$p(\tilde{x}) = \max_{x} p(x)$$

$$n = 8$$

$$\{1,\ 2,\ 3,\ 3,\ 4,\ 4,\ 4,\ 5\}$$

$$\bar{x} = \frac{1}{8}(1+2+3+3+4+4+4+5) = 1 \times \frac{1}{8} + 2 \times \frac{1}{8} + 3 \times \frac{2}{8} + 4 \times \frac{3}{8} + 5 \times \frac{1}{8} = \frac{26}{8} = 3.25$$

3.5

4

X

$$\sigma^2 = Var(x) \triangleq E[(x-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 p(x) dx$$

$$\sigma^2 = Var(x) = E[(x-\mu)^2] = \sum_{i=1}^{n}(x_i-\mu)^2 P_i$$

$$\sigma = \sqrt{Var(x)}$$

$$\sigma^2 = Var(x)$$

$$E[(x-\mu)^2] = E(x^2 - 2\mu x + \mu^2)$$

$$E(x^2) - 2\mu E(x) + \mu^2 = E(x^2) - \mu^2$$

$$\mathcal{N}(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$\int_{-\infty}^{\infty} \mathcal{N}(x, \mu, \sigma) dx = 1$$

$$E(x) = \int_{-\infty}^{\infty} x\mathcal{N}(x, \mu, \sigma)dx = \mu$$

$$Var(x) = \int_{-\infty}^{\infty} (x-\mu)^2 \mathcal{N}(x, \mu, \sigma) dx = \sigma^2$$

$$y = f(x)$$

$p_x(x)$

$$F_y(\psi) = P(y < \psi) = P(f(x) < \psi) = P(x < f^{-1}(\psi)) = \int_{-\infty}^{x} p_X(x)dx$$

$$x \stackrel{\triangle}{=} f^{-1}(y)$$

Y

$$p_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} \int_{-\infty}^{x} p_X(\xi) d\xi = \left[p_X(x) \frac{dx}{dy} \right]_{x=f^{-1}(y)}$$

$$Y = f(X)$$

$$dy/dx > 0$$

$dy/dx < 0$

$$F_Y(y) = P(Y < y) = P(f(X) < y) = P(X > f^{-1}(y)) = \int_x^\infty p_X(\xi) d\xi$$

$$p_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} \int_x^{\infty} p_X(\xi) d\xi = \left[-p_X(x) \frac{dx}{dy} \right]_{x=f^{-1}(y)}$$

$$p_Y(y) = \left[\; p_X(x) \left| \frac{dx}{dy} \right| \; \right]_{x=f^{-1}(y)}$$

$N > 1$

$$\{x_1, \cdots, x_N\}$$

$$\mathbf{x} = [x_1,\ \cdots\ , x_N]^T$$

$$x_i = x(t_i)$$

$x(t)$

x_i

$$\mu_i = E(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \xi_i \, p(\xi_1, \cdots, \xi_N) \, d\xi_1 \cdots d\xi_N$$

$$\mathbf{m} = E(\mathbf{x}) = [E(x_1), \cdots, E(x_N)]^T = [\mu_1, \cdots, \mu_N]^T$$

$$p(x_1, \cdots, x_N)$$

x_j

$$\sigma_{ij}^2$$

$$Cov(x_i, x_j) = E[(x_i - \mu_i)(x_j - \mu_j)] = E(x_i x_j) - E(x_i)\mu_j - \mu_i E(x_j) + \mu_i \mu_j$$

$$E(x_i x_j) - \mu_i \mu_j = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \xi_i \xi_j \, p(\xi_1, \cdots, \xi_N) \, d\xi_1 \cdots d\xi_N - \mu_i \mu_j$$

$$i = j$$

$$\sigma_{ii}^2 = E[(x_i - \mu_i)^2] = E(x_i^2) - \mu_i^2$$

$$\sigma_{ij}^2 = 0$$

$$E(x_i x_j) = E(x_i)E(x_j) = \mu_i \mu_j$$

$$\mathbf{\Sigma}_x = E[(\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})^T] = E(\mathbf{x}\mathbf{x}^T) - \mathbf{m}\mathbf{m}^T = \begin{bmatrix} \sigma_{11}^2 & \cdots & \sigma_{1N}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{N1}^2 & \cdots & \sigma_{NN}^2 \end{bmatrix},$$

$$\sigma_{ii}^2 = E(x_i - \mu_i)^2$$

$$tr\ \mathbf{\Sigma} = \sum_{i=1}^{N} \sigma_i^2 = \sum_{i=1}^{N} \lambda_i$$

$\lambda_1, \cdots, \lambda_N$

$$\Sigma$$

$$\sigma_{ij}^2 = \sigma_{ji}^2$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^T$$

$$\mathbf{z}^T \boldsymbol{\Sigma} \mathbf{z}$$

$$\mathbf{z}^T E[(\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})^T]\mathbf{z} = E[\mathbf{z}^T(\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})^T\mathbf{z}]$$

$$E[(\mathbf{z}^T(\mathbf{x}-\mathbf{m}))^2] \geq 0$$

$$\sigma_{ij} = 0$$

$x_i \quad (i = 1, \cdots, N)$

$$p(x_1, \cdots, x_N) = p(x_1) \cdots p(x_N)$$

$$\sigma_{ij}^2$$

σ_i

σ_j

$$\rho_{ij} = \frac{\sigma_{ij}^2}{\sigma_i \sigma_j}$$

$$\sigma_{ij}^2 = \langle x_i, x_j \rangle$$

$$|\langle x_i,\ x_j\rangle|^2 = \sigma_{ij}^4 \le \langle x_i,\ x_i\rangle\langle x_j,\ x_j\rangle = \sigma_i^2\sigma_j^2,$$

$$\left(\frac{\sigma_{ij}^2}{\sigma_i\sigma_j}\right)^2 = \rho^2 \leq 1$$

$$-1 \leq \rho \leq 1$$

$$\mathbf{R}_x = E(\mathbf{x}\mathbf{x}^T) = \mathbf{\Sigma}_x + \mathbf{m}\mathbf{m}^T = \begin{bmatrix} r_{11}^2 & \cdots & r_{1N}^2 \\ \vdots & \ddots & \vdots \\ r_{N1}^2 & \cdots & r_{NN}^2 \end{bmatrix}$$

$$r_{ij} = E(x_i x_j) = \sigma_{ij}^2 + \mu_i \mu_j \qquad (i, j = 1, \cdots, N)$$

$$\mathbf{m} = 0$$

$$\boldsymbol{\Sigma} = \mathbf{R}$$

$$\mathbf{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_N]$$

$$\mathbf{a}_i^T \mathbf{a}_j = \delta_{ij}$$

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

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \mathbf{A}^T \mathbf{x} = \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_N^T \end{bmatrix} \mathbf{x},$$

$$y_i = \mathbf{a}_i^T \ \mathbf{x}, \qquad (i = 1, \cdots, N)$$

$\mathbf{y}$

$$\mathbf{x} = \mathbf{A}\mathbf{y} = [\mathbf{a}_1, \cdots, \mathbf{a}_N] \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \sum_{i=1}^{N} y_i \mathbf{a}_i$$

$$||\mathbf{x}||^2 = \mathbf{x}^T\mathbf{x} = \left(\sum_{i=1}^{N} y_i \mathbf{a}_i\right)^T \left(\sum_{j=1}^{N} y_j \mathbf{a}_j\right) = \sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \ \mathbf{a}_i^T \mathbf{a}_j = \sum_{i=1}^{N} y_i^2 = ||\mathbf{y}||^2$$

$$\mathbf{m}_y$$

$$\mathbf{\Sigma}_y$$

$\mathbf{m}_x$

$$\mathbf{\Sigma}_x$$

$$\mathbf{m}_y = E(\mathbf{y}) = E(\mathbf{A}^T\mathbf{x}) = \mathbf{A}^T E(\mathbf{x}) = \mathbf{A}^T\mathbf{m}_x$$

$$\mathbf{\Sigma}_y$$

$$E(\mathbf{y}\mathbf{y}^T) - \mathbf{m}_y\mathbf{m}_y^T = E(\mathbf{A}^T\mathbf{x}\mathbf{x}^T\mathbf{A}) - \mathbf{A}^T\mathbf{m}_x\mathbf{m}_x^T\mathbf{A}$$

$$\mathbf{A}^T E(\mathbf{x}\mathbf{x}^T)\mathbf{A} - \mathbf{A}^T \mathbf{m}_x \mathbf{m}_x^T \mathbf{A} = \mathbf{A}^T [E(\mathbf{x}\mathbf{x}^T) - \mathbf{m}_x \mathbf{m}_x^T]\mathbf{A} = \mathbf{A}^T \mathbf{\Sigma}_x \mathbf{A}$$

$$tr(\mathbf{AB}) = tr(\mathbf{BA})$$

$$tr\mathbf{\Sigma_y} = tr(\mathbf{A}^T\mathbf{\Sigma}_x\mathbf{A}) = tr(\mathbf{A}^T\mathbf{A}\mathbf{\Sigma}_x) = tr\mathbf{\Sigma}_x$$

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

m

$$p(\mathbf{x}) = p(x_1, \cdots, x_N)$$

$$\mathbf{x}_k, \quad (k = 1, \cdots, K)$$

$\hat{\mathbf{m}}$

$$\hat{\mathbf{m}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{x}_k$$

$\hat{\Sigma}$

$$\hat{\Sigma}$$

$$\frac{1}{K-1}\sum_{k=1}^{K}(\mathbf{x}_k-\hat{\mathbf{m}})(\mathbf{x}_k-\hat{\mathbf{m}})^T=\frac{1}{K-1}\sum_{k=1}^{K}\left(\mathbf{x}_k\mathbf{x}_k^T-\mathbf{x}_k\hat{\mathbf{m}}^T-\hat{\mathbf{m}}\mathbf{x}_k^T+\hat{\mathbf{m}}\hat{\mathbf{m}}^T\right)$$

$$\frac{1}{K-1}\left(\sum_{k=1}^{K}\mathbf{x}\mathbf{x}^T - K\hat{\mathbf{m}}\hat{\mathbf{m}}^T\right) = \hat{\mathbf{R}} - \frac{K}{K-1}\hat{\mathbf{m}}\hat{\mathbf{m}}^T$$

$$\hat{\mathbf{R}} = \frac{1}{K-1} \sum_{k=1}^{K} \mathbf{x}_k \mathbf{x}_k^T$$

$$K \gg 1$$

$$K - 1 \approx K$$

$$\hat{\boldsymbol{\Sigma}} \approx \frac{1}{K}\sum_{k=1}^{K}\mathbf{x}_k\mathbf{x}_k^T - \hat{\mathbf{m}}\hat{\mathbf{m}}^T, \qquad \hat{\mathbf{R}} \approx \frac{1}{K}\sum_{k=1}^{K}\mathbf{x}_k\mathbf{x}_k^T$$

$$d \times d$$

$K - 1$

K

X

$$\sum_{k=1}^{K}(\mathbf{x}_k - \hat{\mathbf{m}}) = \sum_{k=1}^{K}\mathbf{x}_k - K\hat{\mathbf{m}} = \mathbf{0}$$

$$K - 1 < N$$

$$\mathbf{m} = \mathbf{0}$$

$$\mathbf{y} \neq \mathbf{0}$$

$$\mathbf{y}^T \hat{\mathbf{\Sigma}} \mathbf{y} = \mathbf{y}^T \left(\frac{1}{K} \sum_{k=1}^{K} \mathbf{x}_k \mathbf{x}_k^T \right) \mathbf{y} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{y}^T \mathbf{x}_k \mathbf{x}_k^T \mathbf{y} = \frac{1}{K} \sum_{k=1}^{K} (\mathbf{x}_k^T \mathbf{y})^2 \geq 0$$

$$\mathbf{x}_1^T \mathbf{y} = \cdots = \mathbf{x}_K^T \mathbf{y} = 0$$

$\mathbf{x}_1, \cdots, \mathbf{x}_K$

$$K \geq N$$

$$\mathbf{y}^T \hat{\boldsymbol{\Sigma}} \mathbf{y} = \frac{1}{K} \sum_{k=1}^{K} (\mathbf{x}_k^T \mathbf{y})^2 = 0$$

$$\mathbf{y} = \sum_{i=1}^{K} a_i \mathbf{x}_i$$

$$\mathbf{y}^T\mathbf{y} = \sum_{i=1}^{K} a_i \mathbf{x}_i^T \mathbf{y} = 0$$

$$\mathbf{y}^T \hat{\mathbf{\Sigma}} \mathbf{y} = 0$$

$$\mathbf{y} = \mathbf{0}$$

$$\mathbf{y}^T \hat{\mathbf{\Sigma}} \mathbf{y} \neq 0$$

$\lambda_i > 0 \ (i = 1, \cdots, N)$

$$tr\ \mathbf{\Sigma} = \sum_{i=1}^{N} \lambda_i > 0, \qquad \det\ \mathbf{\Sigma} = \prod_{i=1}^{N} \lambda_i > 0$$

$$y = cx$$

c

μ_y

$$E(y) = E(cx) = c\,E(x) = c\,\mu_x$$

σ_y

$$E(y^2) - \mu_y^2 = E[(cx)^2] - (c\mu_x)^2 = c^2\,E(x^2) - c^2\mu_x^2 = c^2(E(x^2) - \mu_x^2) = \sigma_x^2$$

$$y = x_1 + \cdots + x_n$$

$$\mu_y = E(y) = E\left(\sum_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} E(x_i) = \sum_{i=1}^{n} \mu_{x_i}$$

$$E(y^2) = E\left(\sum_{i=1}^{n} x_i\right)^2 = E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} E(x_i x_j)$$

$$(E(y))^2 = \left(\sum_{i=1}^{n} E(x_i)\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} E(x_i)E(x_j)$$

$$Var(y) = E(y^2) - (E(y))^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} E(x_i x_j) - \sum_{i=1}^{n}\sum_{j=1}^{n} E(x_i)E(x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(x_i, x_j)$$

$$Cov(x_i, x_j) = 0$$

$$i \neq j$$

$$\sigma_y^2 = Var(y) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(x_i, x_j) = \sum_{i=1}^{n} Cov(x_i, x_i) = \sum_{i=1}^{n} Var(x_i) = \sum_{i=1}^{n} \sigma_{x_i}^2$$

μ_x

$$y = f(x) = f(\mu_x) + f'(\mu_x)(x - \mu_x) + \frac{f''(\mu_x)}{2}(x - \mu_x)^2 + \cdots \approx f(\mu_x) + f'(\mu_x)(x - \mu_x)$$

$$|x - \mu_x|$$

$$\mu_y = E(y) \approx E[f(\mu_x) + f'(\mu_x)(x - \mu_x)] = f(\mu_x)$$

$$Var(y) \approx Var[f(\mu_x) + f'(\mu_x)(x - \mu_x)] = Var[f'(\mu_x)(x - \mu_x)] = f'^2(\mu_x)Var(x) = f'^2(\mu_x)\sigma_x^2$$

$$y = f(x_1, \cdots, x_n) = f(\mathbf{x})$$

$$\mathbf{x} = [x_1, \cdots, x_n]^2$$

$$\mathbf{m}_x = [\mu_{x_1}, \cdots, \mu_{x_n}]^T$$

$$f(x_1, \cdots, x_n)$$

$$f(\mu_{x_1}, \cdots, \mu_{x_n}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_i - \mu_{x_i}) + \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial f^2}{\partial x_i \partial x_j}(x_i - \mu_{x_i})(x_j - \mu_{x_j}) + \cdots$$

$\approx$

$$f(\mu_{x_1}, \cdots, \mu_{x_n}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_i - \mu_{x_i})$$

$$f(\mathbf{x}) = f(\mathbf{m}_x) + \mathbf{g}(\mathbf{m}_\mathbf{x})^T(\mathbf{x} - \mathbf{m}_x) + \frac{1}{2}(\mathbf{x} - \mathbf{m}_x)^T\mathbf{H}(\mathbf{m}_x)(\mathbf{x} - \mathbf{m}_x) + \cdots \approx f(\mathbf{m}_x) + \mathbf{g}(\mathbf{m}_\mathbf{x})^T(\mathbf{x} - \mathbf{m}_x)$$

$$\mathbf{g}(\mathbf{m}_x) = \begin{bmatrix} \frac{\partial f \mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f \mathbf{x})}{\partial x_n} \end{bmatrix}_{\mathbf{m}_x} , \quad \mathbf{H}(\mathbf{m}_x) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{bmatrix}_{\mathbf{m}_x}$$

$$f(\mathbf{x})$$

$$Ef(\mathbf{x}) \approx E\left[f(\mu_{x_1}, \cdots, \mu_{x_n}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_i - \mu_{x_i})\right]$$

$$f(\mu_{x_1}, \cdots, \mu_{x_n}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} E(x_i - \mu_{x_i}) = f(\mu_{x_1}, \cdots, \mu_{x_n})$$

σ_y^2

$$Var\left[f((x_1,\cdots,x_n)\right] \approx Var\left[f(\mu_{x_1},\cdots,\mu_{x_n}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_i - \mu_{x_i})\right]$$

$$Var\left[\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(x_i-\mu_{x_i})\right]=\sum_{i=1}^{n}\left(\frac{\partial f}{\partial x_i}\right)^2\sigma_{x_i}^2$$

$$\sigma_y = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_{x_i}^2}$$

$$p(x) = \mathcal{N}(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$\mathbf{x} = [x_1, \cdots, x_n]^T$$

$$p(x_1, \cdots, x_n)$$

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}, \mathbf{m}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2} \left|\boldsymbol{\Sigma}\right|^{1/2}} \; \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{m})\right]$$

$$\exp\left(-\frac{1}{2}(n\log(2\pi)+\log|\mathbf{\Sigma}|)\right)\exp\left[-\frac{1}{2}(\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x}-2\mathbf{m}^T\mathbf{\Sigma}^{-1}\mathbf{x}+\mathbf{m}^T\mathbf{\Sigma}^{-1}\mathbf{m})\right]$$

$$\exp\left(-\frac{1}{2}(n\log(2\pi)+\log|\mathbf{\Sigma}|+\mathbf{m}^T\mathbf{\Sigma}^{-1}\mathbf{m})\right)\exp\left[-\frac{1}{2}(\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x}-2\mathbf{m}^T\mathbf{\Sigma}^{-1}\mathbf{x})\right]$$

$$\mathbf{m} = E(\mathbf{x})$$

$$\mathbf{\Sigma} = Cov(\mathbf{x})$$

$$\psi(\mathbf{x}) = \log \mathcal{N}(\mathbf{x}, \mathbf{m}, \boldsymbol{\Sigma}) = -\frac{1}{2}(\mathbf{x} - \mathbf{m})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \mathbf{m}) - \frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}|$$

$$\mathbf{g}_{\psi}(\mathbf{x})$$

$$\frac{d}{d\mathbf{x}}\psi(\mathbf{x}) = -\mathbf{\Sigma}^{-1}(\mathbf{x} - \mathbf{m})$$

$$\mathbf{H}_{\psi}(\mathbf{x})$$

$$\frac{d^2}{d\mathbf{x}^2}\psi(\mathbf{x}) = -\boldsymbol{\Sigma}^{-1}$$

$$\mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$$

$$\mathbf{g}_{\psi}(\mathbf{x}) = -\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \mathbf{m}) = \mathbf{0}$$

$$\mathbf{x} = \mathbf{m}$$

$$\boldsymbol{\Sigma} = -\mathbf{H}_{\psi}^{-1}(\mathbf{x})$$

$$\mathcal{N}(\mathbf{x}, \mathbf{m}, \mathbf{\Sigma}) = c_0$$

c_0

$$(\mathbf{x}-\mathbf{m})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\mathbf{m}) = c_1$$

c_1

$n = 2$

$$\mathbf{x} = [x_1,\ x_2]^T$$

$$\mathbf{\Sigma}^{-1} = \begin{bmatrix} A & B/2 \\ B/2 & C \end{bmatrix}$$

$$(\mathbf{x}-\mathbf{m})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{m}) = [x_1-\mu_1, x_2-\mu_2]\begin{bmatrix} A & B/2 \\ B/2 & C \end{bmatrix}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}$$

$$A(x_1 - \mu_1)^2 + B(x_1 - \mu_1)(x_2 - \mu_2) + C(x_2 - \mu_2)^2$$

$$Ax_1^2 + Bx_1x_2 + Cx_2^2 - (2A\mu_1 + B\mu_2)x_1 - (2C\mu_2 + B\mu_1)x_2 + (A\mu_1^2 + B\mu_1\mu_2 + C\mu_2^2) = c_1$$

$$\boldsymbol{\Sigma}^{-1}$$

$$\left|\mathbf{\Sigma}^{-1}\right| = AC - B^2/4 \geq 0,$$

$$\Delta = B^2 - 4AC < 0$$

$$\mathbf{m} = [\mu_1, \mu_2]^T$$

$$\Delta = B^2 - 4AC$$

$$f(x, y) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$$

$$f(x, y)$$

$$\begin{cases} \text{an ellipse} & \text{if } \Delta < 0 \\ \text{a parabola} & \text{if } \Delta = 0 \\ \text{a hyperbola} & \text{if } \Delta > 0 \end{cases}$$

$n = 3$

$n > 3$

$$\mathcal{N}(\mathbf{x}, \mathbf{m}, \mathbf{\Sigma}) = c_0$$

$$\mathbf{\Sigma} = diag[\sigma_1^2, \cdots, \sigma_n^2] = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

$$(\mathbf{x}-\mathbf{m})^T \mathbf{\Sigma}^{-1} (\mathbf{x}-\mathbf{m}) = \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2} = c_1$$

$$p(\mathbf{x}_i) = \mathcal{N}(\mathbf{m}_i, \boldsymbol{\Sigma}_i) \ (i = 1, 2)$$

$$p(\mathbf{x}_1 + \mathbf{x}_2) = \mathcal{N}(\mathbf{m}_1 + \mathbf{m}_2,\, \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)$$

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{m}_x, \mathbf{\Sigma}_x)$$

$$p(\mathbf{y}) = p(\mathbf{A}\mathbf{x} + \mathbf{b}) = \mathcal{N}(\mathbf{m}_y, \mathbf{\Sigma}_y)$$

$$\mathbf{m}_y = \mathbf{A}\mathbf{m}_x + \mathbf{b}, \qquad \mathbf{\Sigma}_y = \mathbf{A}\mathbf{\Sigma}_x\mathbf{A}^T$$

$$\mathbf{m}_y = E(\mathbf{A}\mathbf{x} + \mathbf{b}) = \mathbf{A}\, E(\mathbf{x}) + \mathbf{b} = \mathbf{A}\mathbf{m}_x + \mathbf{b}$$

$$E[(\mathbf{y} - \mathbf{m}_y)(\mathbf{y} - \mathbf{m}_y)^T]$$

$$E[[(\mathbf{A}\mathbf{x}-\mathbf{b})-(\mathbf{A}\mathbf{m}_x-\mathbf{b})][(\mathbf{A}\mathbf{x}-\mathbf{b})-(\mathbf{A}\mathbf{m}_x+\mathbf{b})]^T]$$

$$E[(\mathbf{Ax}-\mathbf{Am}_x)(\mathbf{Ax}-\mathbf{Am}_x)^T] = \mathbf{A}E[(\mathbf{x}-\mathbf{m}_x)(\mathbf{x}-\mathbf{m}_x)^T]\mathbf{A}^T$$

$$\mathbf{A}\boldsymbol{\Sigma}_x\mathbf{A}^T$$

$$\mathbf{A} = \mathbf{I}$$

$$p(\mathbf{x}+\mathbf{b}) = \mathcal{N}(\mathbf{m}_x + \mathbf{b}, \boldsymbol{\Sigma}_x)$$

$$\mathbf{b} = \mathbf{0}$$

$$p(\mathbf{A}\mathbf{x}) = \mathcal{N}(\mathbf{m}_x, \mathbf{A}\boldsymbol{\Sigma}_x\mathbf{A}^T)$$

$$p(\mathbf{x}) = N\left([\mu_1, \cdots, \mu_n]^T, \, diag[\sigma_1^2, \cdots, \sigma_n^2]\right)$$

$$p(\sum_{i=1}^{n} x_i) = \mathcal{N}(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)$$

$$p(x_1, x_2) = p(\mathbf{x}) = p\left(\left[\begin{array}{c} x_1 \\ x_2 \end{array}\right]\right) = N\left(\left[\begin{array}{c} \mu_1 \\ \mu_2 \end{array}\right], \left[\begin{array}{cc} \sigma_1^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_2^2 \end{array}\right]\right)$$

$$\mathbf{A}_1 = [1,\ 0]$$

$$\mathbf{A}_2 = [0,\ 1]$$

$$\mathbf{A}_1\mathbf{x} = x_1$$

$$\mathbf{A}_2\mathbf{x} = x_2$$

$$p(x_1) = \mathcal{N}(\mu_1, \sigma_1^2), \qquad p(x_1) = \mathcal{N}(\mu_2, \sigma_2^2)$$

$p_i(\mathbf{x})$

$$\mathcal{N}(\mathbf{x}, \mathbf{m}_i, \mathbf{\Sigma}_i) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}_i|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{m}_i)^T \mathbf{\Sigma}_i^{-1}(\mathbf{x} - \mathbf{m}_i)\right]$$

$$\frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}_i|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}^T\mathbf{\Sigma}_i^{-1}\mathbf{x} - 2\mathbf{x}^T\mathbf{\Sigma}_i^{-1}\mathbf{m}_i + \mathbf{m}_i\mathbf{\Sigma}_i^{-1}\mathbf{m}_i)\right]$$

$$\exp\left[-\frac{1}{2}[n\log(2\pi)+\log|\mathbf{\Sigma}_i|+\mathbf{m}_i\mathbf{\Sigma}_i^{-1}\mathbf{m}_i]\right]\exp\left[-\frac{1}{2}(\mathbf{x}^T\mathbf{\Sigma}_i^{-1}\mathbf{x}-2\mathbf{x}^T\mathbf{\Sigma}_i^{-1}\mathbf{m}_i)\right]$$

$$\exp\left[-\frac{1}{2}[\mathbf{x}^T\boldsymbol{\Sigma}_i^{-1}\mathbf{x} - 2\mathbf{x}^T\boldsymbol{\Sigma}_i^{-1}\mathbf{m}_i + c_i]\right] \qquad (i = 1, 2)$$

$$c_i = n \log(2\pi) + \log|\mathbf{\Sigma}_i| + \mathbf{m}_i \mathbf{\Sigma}_i^{-1} \mathbf{m}_i$$

$$p_1(\mathbf{x})p_2(\mathbf{x})$$

$$\mathcal{N}(\mathbf{x}, \mathbf{m}_1, \mathbf{\Sigma}_2)\mathcal{N}(\mathbf{x}, \mathbf{m}_2, \mathbf{\Sigma}_2)$$

$$\exp\left[-\frac{1}{2}[\mathbf{x}^T\mathbf{\Sigma}_1^{-1}\mathbf{x} - 2\mathbf{x}^T\mathbf{\Sigma}_1^{-1}\mathbf{m}_1 + c_1 + \mathbf{x}^T\mathbf{\Sigma}_2^{-1}\mathbf{x} - 2\mathbf{x}^T\mathbf{\Sigma}_2^{-1}\mathbf{m}_2 + c_2]\right]$$

$$\exp\left[-\frac{1}{2}[\mathbf{x}^T(\mathbf{\Sigma}_1^{-1}+\mathbf{\Sigma}_2^{-1})\mathbf{x}-2\mathbf{x}^T(\mathbf{\Sigma}_1^{-1}\mathbf{m}_1+\mathbf{\Sigma}_2^{-1}\mathbf{m}_2)+c_1+c_2]\right]$$

$$\exp\left[-\frac{1}{2}[\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x} - 2\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{m} + c + (c_1 + c_2 - c)]\right]$$

$$\mathcal{N}(\mathbf{x}, \mathbf{m}, \mathbf{\Sigma})\ \exp(c_1 + c_2 - c) \propto \mathcal{N}(\mathbf{x}, \mathbf{m}, \mathbf{\Sigma})$$

$$c = n \log(2\pi) + \log|\mathbf{\Sigma}| + \mathbf{m}^T \mathbf{\Sigma}^{-1} \mathbf{m}$$

$$\mathbf{\Sigma} = (\mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1})^{-1}$$

$$\mathbf{\Sigma}^{-1}\mathbf{m} = \mathbf{\Sigma}_1^{-1}\mathbf{m}_1 + \mathbf{\Sigma}_2^{-1}\mathbf{m}_2$$

$$\mathbf{m} = \boldsymbol{\Sigma}(\boldsymbol{\Sigma}_1^{-1}\mathbf{m}_1 + \boldsymbol{\Sigma}_2^{-1}\mathbf{m}_2) = (\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1})^{-1}(\boldsymbol{\Sigma}_1^{-1}\mathbf{m}_1 + \boldsymbol{\Sigma}_2^{-1}\mathbf{m}_2)$$

$$\mathbf{x} = [\mathbf{x}_!^T, \mathbf{x}_2^T]^T$$

$$\mathcal{N}(\mathbf{x}, \mathbf{m}, \mathbf{\Sigma})$$

$\mathbf{x}_1$

$$\mathbf{x}_2$$

$$p\left(\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix}\right) = p(\mathbf{x}) = \mathcal{N}(\mathbf{x}, \mathbf{m}, \boldsymbol{\Sigma}) = N\left(\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix}, \begin{bmatrix}\mathbf{m}_1\\ \mathbf{m}_2\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\right)$$

$$\mathbf{\Sigma}_{ij} = \mathbf{\Sigma}_{ji}^{T}$$

$$\boldsymbol{\Sigma}_{ii} = \boldsymbol{\Sigma}_{ii}^{T}, \ (i, j = 1, 2, \ i \neq j)$$

$\mathbf{x}_i$

$$\mathbf{x}_j$$

$$p(\mathbf{x}_i/\mathbf{x}_j) = \mathcal{N}(\mathbf{x}_i, \mathbf{m}_{i/j}, \mathbf{\Sigma}_{i/j})$$

$$\left\{ \begin{array}{lcl} \mathbf{m}_{i/j} & = & \mathbf{m}_i + \mathbf{\Sigma}_{ij}\mathbf{\Sigma}_{jj}^{-1}(\mathbf{x}_j - \mathbf{m}_j) \\ \mathbf{\Sigma}_{i/j} & = & \mathbf{\Sigma}_{ii} - \mathbf{\Sigma}_{ij}\mathbf{\Sigma}_{jj}^{-1}\mathbf{\Sigma}_{ji} \end{array} \right.$$

$p(\mathbf{x})$

$$p(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{m})\right]$$

$$\frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}} \ \exp\left[-\frac{1}{2}Q(\mathbf{x}_1, \mathbf{x}_2)\right]$$

$$Q(\mathbf{x}_1, \mathbf{x}_2)$$

$$(\mathbf{x}-\mathbf{m})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{m})$$

$$[(\mathbf{x}_1 - \mathbf{m}_1)^T, (\mathbf{x} - \mathbf{m}_2)^T] \begin{bmatrix} \mathbf{\Sigma}^{11} & \mathbf{\Sigma}^{12} \\ \mathbf{\Sigma}^{21} & \mathbf{\Sigma}^{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 - \mathbf{m}_1 \\ \mathbf{x}_2 - \mathbf{m}_2 \end{bmatrix}$$

$$(\mathbf{x}_1 - \mathbf{m}_1)^T \boldsymbol{\Sigma}^{11} (\mathbf{x}_1 - \mathbf{m}_1) + 2(\mathbf{x}_1 - \mathbf{m}_1)^T \boldsymbol{\Sigma}^{12} (\mathbf{x}_2 - \mathbf{m}_2) + (\mathbf{x}_2 - \mathbf{m}_2)^T \Sigma^{22} (\mathbf{x}_2 - \mathbf{m}_2)$$

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}^{11} & \boldsymbol{\Sigma}^{12} \\ \boldsymbol{\Sigma}^{21} & \boldsymbol{\Sigma}^{22} \end{bmatrix}$$

$$\mathbf{\Sigma}^{11}$$

$$(\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})^{-1} = \mathbf{\Sigma}_{11}^{-1} + \mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}$$

$\boldsymbol{\Sigma}^{22}$

$$(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1} = \boldsymbol{\Sigma}_{22}^{-1} + \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$$

$$\Sigma^{12}$$

$$-\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}=(\boldsymbol{\Sigma}^{12})$$

$$\boldsymbol{\Sigma}^{11}$$

$\boldsymbol{\Sigma}^{22}$

Σ^{12}

$$Q(\mathbf{x}_1, \mathbf{x}_2)$$

$$(\mathbf{x}_1 - \mathbf{m}_1)^T [\mathbf{\Sigma}_{11}^{-1} + \mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12} (\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12})^{-1} \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{11}^{-1}] (\mathbf{x}_1 - \mathbf{m}_1)$$

$$-2(\mathbf{x}_1 - \mathbf{m}_1)^T [\mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12} (\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12})^{-1}] (\mathbf{x}_2 - \mathbf{m}_2)$$

$$+(\mathbf{x}_2 - \mathbf{m}_2)^T[(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}](\mathbf{x}_2 - \mathbf{m}_2)$$

$$(\mathbf{x}_1 - \mathbf{m}_1)^T \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{x}_1 - \mathbf{m}_1)$$

$$+(\mathbf{x}_1-\mathbf{m}_1)^T\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}(\mathbf{\Sigma}_{22}-\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}(\mathbf{x}_1-\mathbf{m}_1)$$

$$-2(\mathbf{x}_1 - \mathbf{m}_1)^T \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} (\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12})^{-1} (\mathbf{x}_2 - \mathbf{m}_2)$$

$$+(\mathbf{x}_2 - \mathbf{m}_2)^T(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}(\mathbf{x}_2 - \mathbf{m}_2)$$

$$+[(\mathbf{x}_2 - \mathbf{m}_2) - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \mathbf{m}_1)]^T(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}[(\mathbf{x}_2 - \mathbf{m}_2) - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \mathbf{m}_1)]$$

u

v

$$A = A^T$$

$$u^T A u - 2u^T A v + v^T A v = u^T A u - u^T A v - u^T A v + v^T A v$$

$$u^T A(u-v) - (u-v)^T A v = u^T A(u-v) - v^T A(u-v)$$

$$(u-v)^T A(u-v) = (v-u)^T A(v-u)$$

$$\begin{cases} \mathbf{b} = \mathbf{m}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \mathbf{m}_1) \\ \mathbf{A} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \end{cases}$$

$$\left\{\begin{array}{lll} Q_1(\mathbf{x}_1) & = & (\mathbf{x}_1-\mathbf{m}_1)^T\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1-\mathbf{m}_1) \\ Q_2(\mathbf{x}_1,\mathbf{x}_2) & = & [(\mathbf{x}_2-\mathbf{m}_2)-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1-\mathbf{m}_1)]^T(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}[(\mathbf{x}_2-\mathbf{m}_2)-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1-\mathbf{m}_1)] \\ & = & (\mathbf{x}_2-\mathbf{b})^T\mathbf{A}^{-1}(\mathbf{x}_2-\mathbf{b}) \end{array}\right.$$

$$Q(\mathbf{x}_1, \mathbf{x}_2) = Q_1(\mathbf{x}_1) + Q_2(\mathbf{x}_1, \mathbf{x}_2)$$

$$p(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}Q(\mathbf{x}_1, \mathbf{x}_2)\right]$$

$$\frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}_{11}|^{1/2}|\mathbf{\Sigma}_{22}-\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}|^{1/2}}\ \exp\left[-\frac{1}{2}Q_1(\mathbf{x}_1)\right]\exp\left[-\frac{1}{2}Q_2(\mathbf{x}_1,\mathbf{x}_2)\right]$$

$$\frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}_{11}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}_1 - \mathbf{m}_1)^T \mathbf{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \mathbf{m}_1)\right] \frac{1}{(2\pi)^{q/2}|\mathbf{A}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}_2 - \mathbf{b})^T \mathbf{A}^{-1}(\mathbf{x}_2 - \mathbf{b})\right]$$

$$\mathcal{N}(\mathbf{x}_1, \mathbf{m}_1, \mathbf{\Sigma}_{11})\,\mathcal{N}(\mathbf{x}_2, \mathbf{b}, \mathbf{A})$$

$$|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_{11}| \; |\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}|$$

$p_1(\mathbf{x}_1)$

$$\int p(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x_2} = \mathcal{N}(\mathbf{x}_1, \mathbf{m}_1, \mathbf{\Sigma}_{11}) \int \mathcal{N}(\mathbf{x}_2, \mathbf{b}, \mathbf{A}) d\mathbf{x_2}$$

$$\mathcal{N}(\mathbf{x}_1, \mathbf{m}_1, \mathbf{\Sigma}_{11}) = \frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}_{11}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}_1 - \mathbf{m}_1)^T \mathbf{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \mathbf{m}_1)\right]$$

$$p_{2|1}(\mathbf{x}_2|\mathbf{x}_1) = \frac{p(\mathbf{x}_1, \mathbf{x}_2)}{p_1(\mathbf{x}_1)} = \mathcal{N}(\mathbf{x}_2, \mathbf{b}, \mathbf{A}) = \frac{1}{(2\pi)^{q/2}|\mathbf{A}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}_2 - \mathbf{b})^T \mathbf{A}^{-1}(\mathbf{x}_2 - \mathbf{b})\right]$$

$$\left\{ \begin{array}{lll} \mathbf{b} & = & \mathbf{m}_{2/1} = \mathbf{m}_2 + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \mathbf{m}_1) \\ \mathbf{A} & = & \mathbf{\Sigma}_{2/1} = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12} \end{array} \right.$$

$$p(x) = \mathcal{N}(0, 1)$$

$$E(x) = 0$$

$$Var(x) = E(x^2) = 1$$

$(0,\ 1)$

$$(-\infty,\ \infty)$$

$$p(x) = \mathcal{N}(0,\ 1)$$

$$F(x) = \int_{-\infty}^{x} p(u)\, du$$

$$x' = \mu + \sigma x$$

$$p(x') = \mathcal{N}(\mu,\, \sigma^2)$$

$E(x')$

$$E(\mu + \sigma x) = \mu + \sigma E(x) = \mu$$

$$Var(x')$$

$$[(x' - \mu)^2] = E(\sigma^2 x) = E(\sigma^2 x^2) = \sigma^2 E(x^2) = \sigma^2$$

$x_1, \cdots, x_n$

$$E(x_i) = 0$$

$$Var(x_i) = 1$$

$$Cov(x_i, x_j) = \delta_{ij}$$

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{0},\, \mathbf{I})$$

$$\mathbf{m}_x = \mathbf{0}$$

$$\mathbf{\Sigma}_x = \mathbf{I}$$

$N(\mathbf{m},\ \mathbf{\Sigma})$

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{m}$$

$\mathbf{m}_y$

$$E(\mathbf{y}) = \mathbf{W}\mathbf{m}_x + \mathbf{m} = \mathbf{m}$$

$$\mathbf{W}\boldsymbol{\Sigma}_x\mathbf{W}^T = \mathbf{W}\mathbf{W}^T$$

$$\mathbf{m}_y = \mathbf{m}$$

$$\mathbf{\Sigma}_y = \mathbf{W}\mathbf{W}^T = \mathbf{\Sigma}$$

W

L

$$\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T$$

$$\mathbf{\Sigma} = \mathbf{U}\mathbf{S}\mathbf{V}^T$$

$$\boldsymbol{\Sigma}^T = \boldsymbol{\Sigma}$$

$$\mathbf{U} = \mathbf{V}$$

S

$$\mathbf{W} = \mathbf{U}\mathbf{S}^{1/2}$$

$$\mathbf{W}\mathbf{W}^T = (\mathbf{U}\mathbf{S}^{1/2})\,(\mathbf{U}\mathbf{S}^{1/2})^T = \mathbf{U}\mathbf{S}\mathbf{U}^T = \boldsymbol{\Sigma}$$

$$\mathbf{L}\mathbf{L}^T = \mathbf{W}\mathbf{W}^T = \mathbf{\Sigma}$$

$$\mathbf{L} \neq \mathbf{W}$$

L

$$\mathbf{y} = \mathbf{L}\mathbf{x} + \mathbf{m}$$

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{m},\ \boldsymbol{\Sigma})$$

$$P_i = P(x_i)$$

$$\sum_{i=1}^{N} P_i = 1$$

$N = 26$

$$P(x = 'a') > P(x = 'z')$$

P_i

p_i

$$\log_2(1/P_i) = -\log_2 P_i$$

$$H(P) = E\left(\log \frac{1}{P_i}\right) = \sum_i P_i \left(\log \frac{1}{P_i}\right) = -\sum_i P_i \log_2 P_i$$

H

$$N = 2^n$$

$$p_i = 1/N \ (i = 1, \cdots, N)$$

$$\frac{\partial}{\partial P_j}\left[-\sum_{i=1}^{N} P_i log\ P_i + \lambda\left(1 - \sum_{i=0}^{N} P_i\right)\right] = 0$$

$$-\frac{\partial}{\partial P_j} P_j \ log \ P_j - \lambda = -(log \ P_j + 1) - \lambda = 0$$

$$P_j = 2^{-\lambda-1} \qquad (j = 1, \cdots, N)$$

$$\sum_{j=1}^{N} P_j = \sum_{j=1}^{N} 2^{-\lambda-1} = 2^{-\lambda-1} N = 1$$

λ

$$\lambda = \log N - 1$$

P_j

$$P_j = 2^{-\lambda-1} = 2^{-\log N} = \frac{1}{N}, \qquad (j = 1, \cdots, N)$$

$$H_{max}(E_1, \cdots, E_N) = -\sum_{i=1}^{N} \frac{1}{N} log_2 \frac{1}{N} = log_2 N = n$$

$$\log_2 \frac{1}{P_i} = \log_2 \left(\frac{1}{1/2^n} \right) = \log_2 2^n = n$$

$$-\sum_{i=1}^{N} P_i \log_2 P = -\sum_{i=1}^{N} \frac{1}{N} \log_2 \frac{1}{2^n} = \sum_{i=1}^{N} \frac{1}{N} \log_2 2^n = n$$

$$N = 2^n = 2^3 = 8$$

$$P_i = 1/2^3 = 1/8$$

$N = 3$

$N = 2$

$$P_1 = 1$$

$$P_2 = 0$$

$$x \log_2 x|_{x=0} = 0$$

$$H = -1 \ log_2 \ 1 - 0 \ log_2 \ 0 = 0$$

$$P_1 = 0.1$$

$$P_2 = 0.9$$

$$H = -(0.1\ log_2\ 0.1 + 0.9\ log_2\ 0.9) = 0.47$$

$$P_1 = 0.2$$

$$P_2 = 0.8$$

$$H = -(0.2\ log_2\ 0.2 + 0.8\ log_2\ 0.8) = 0.72$$

$$P_1 = 0.3$$

$$P_2 = 0.7$$

$$H = -(0.3\ log_2\ 0.3 + 0.7\ log_2\ 0.7) = 0.88$$

$$P_1 = 0.4$$

$$P_2 = 0.6$$

$$H = -(0.4\ log_2\ 0.4 + 0.6\ log_2\ 0.6) = 0.97$$

$$P_1 = P_2 = 0.5 = 1/2$$

$H = 1$

$$H = -\frac{1}{2}\ log_2\ \frac{1}{2} - \frac{1}{2}\ log_2\ \frac{1}{2} = \frac{1}{2} + \frac{1}{2} = 1$$

$$H(p) = -\int p(x) \log p(x)\ dx$$

$$2^{H(p)}$$

$$x \in [a,\ b]$$

$H(p)$

$$L(p) = H(p) + \lambda \left(1 - \int_a^b p(x)\, dx \right) = \int_a^b \left[-p(x)\ \log p(x) - \lambda p(x) \right]\ dx + \lambda$$

$$\frac{d\,L(p)}{d\,p(x)} = -(\log\, p(x) + 1) - \lambda = 0$$

$$p(x) = e^{-\lambda - 1}$$

$$\int_{-\infty}^{\infty} p(x)\, dx = \int_a^b e^{-\lambda-1}\, dx = e^{-\lambda-1}(b-a) = 1$$

$$p(x) = e^{-\lambda-1} = \frac{1}{b-a}$$

$[a,\ b]$

$$H(p) = -\int_a^b \frac{1}{b-a} \log\left(\frac{1}{b-a}\right) = \log(b-a)$$

$$x \in (-\infty,\ \infty)$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 p(x)\, dx$$

$$\sigma^2$$

$$L(p) = -\int_{-\infty}^{\infty} p(x)\, \log p(x)\, dx + \lambda_1 \left(1 - \int_{-\infty}^{\infty} p(x)\, dx\right) + \lambda_2 \left(\sigma^2 - \int_{-\infty}^{\infty} (x-\mu)^2\, p(x)\, dx\right)$$

$$\frac{d\,L(p)}{d\,p(x)} = -(log\,p(x) + 1) - \lambda_1 - \lambda_2(x - \mu)^2 = 0$$

$$p(x) = e^{-(\lambda_1 + \lambda_2 (x-\mu)^2 + 1)} = e^{-(\lambda_1 + 1)} e^{-\lambda_2 (x-\mu)^2}$$

$$1 = \int_{-\infty}^{\infty} p(x)\, dx = e^{-\lambda_1 - 1} \int_{-\infty}^{\infty} e^{-\lambda_2 (x-\mu)^2} dx = e^{-(\lambda_1 + 1)} \sqrt{\frac{\pi}{\lambda_2}}$$

$$\sigma^2$$

$$\int_{-\infty}^{\infty} (x-\mu)^2 p(x)\, dx = e^{-\lambda_1 - 1} \int_{-\infty}^{\infty} e^{-\lambda_2 (x-\mu)^2} (x-\mu)^2\, dx$$

$$e^{-\lambda_1 - 1} \frac{1}{2} \sqrt{\frac{\pi}{\lambda_2^3}} = e^{-\lambda_1 - 1} \frac{1}{2\lambda_2} \sqrt{\frac{\pi}{\lambda_2}}$$

$$\lambda_2 = \frac{1}{2\sigma^2}$$

$$e^{-(\lambda_1+1)}\sqrt{2\pi\sigma^2} = 1,$$

$$e^{-(\lambda_1+1)} = \frac{1}{\sqrt{2\pi\sigma^2}}$$

$$\lambda_2 = 1/2\sigma^2$$

$$p(x) = e^{-(\lambda_1+1)} e^{-\lambda_2 (x-\mu)^2} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \mathcal{N}(x, \mu, \sigma^2)$$

$H(g)$

$$-\int_{-\infty}^{\infty} g(x)\log g(x)\,dx = -\int_{-\infty}^{\infty} g(x)\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/2\sigma^2}\right)\;dx$$

$$-\int_{-\infty}^{\infty} g(x) \left(log \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{(x-\mu)^2}{2\sigma^2} \right) dx$$

$$\frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\int_{-\infty}^{\infty}(x-\mu)^2 g(x)\,dx$$

$$\frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sigma^2 = \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2}\log(e)$$

$$\frac{1}{2}\log(2\pi e\sigma^2) = \frac{1}{2}\log(2\pi e) + \log\sigma$$

$$\sigma^2 = \int (x - \mu)^2 p(x)\, dx$$

$$H = -\int p(x)\ \log\ p(x)\ dx$$

$$H(x,y) = -\sum_i \sum_j P(x_i, y_j) \log P(x_i, y_j)$$

$$H(x_1, \cdots, x_n) = -\sum_{x_1} \cdots \sum_{x_n} P(x_1, \cdots, x_n) \log P(x_1, \cdots, x_n)$$

$$H(y|x)$$

$$H(y|x = x_i)$$

$$x = x_i$$

$$H(y|x = x_i) = -\sum_j P(y_j|x_i) \log P(y_j|x_i)$$

$$H(y|x)$$

$$E_x\left(H(y|x=x_i)\right)=\sum_i P(x_i)H(y|x=x_i)$$

$$-\sum_i P(x_i) \sum_j P(y_j|x_i) \log P(y_j|x_i) = -\sum_i \sum_j P(x_i) P(y_j|x_i) \log P(y_j|x_i)$$

$$-\sum_i \sum_j P(x_i, y_j) \log P(y_j|x_i)$$

$$P(x, y) = P(y|x)P(x)$$

$$-\sum_i \sum_j P(x_i, y_j) \log P(y_j|x_i) = \sum_i \sum_j P(x_i, y_j) \log \frac{P(x_i)}{P(x_i, y_j)}$$

$$-\sum_i \sum_j P(x_i, y_j) \log P(x_i, y_j) + \sum_i \sum_j P(x_i, y_j) \log P(x_i)$$

$$H(x,y) + \sum_i P(x_i) \log P(x_i) = H(x,y) - H(x)$$

$$H(x|y) = H(x,y) - H(y)$$

$$H(x,y) = H(y|x) + H(x) = H(x|y) + H(y)$$

$$H(y|x) = H(x|y) + H(y) - H(x)$$

$I(x, y)$

$$I(x, y)$$

$$\sum_i \sum_j P(x_i, y_j) \log \frac{P(x_i, y_j)}{P(x_i)P(y_j)}$$

$$\sum_i \sum_j P(x_i, y_j) \log \frac{P(x_i, y_j)}{P(x_i)} - \sum_i \sum_j P(x_i, y_j) \log P(y_j)$$

$$\sum_i \sum_j P(y_j|x_i)P(x_i) \log P(y_j|x_i) - \sum_j \log P(y_j) \left(\sum_i P(x_i, y_j) \right)$$

$$\sum_i P(x_i) \sum_j P(y_j|x_i) \log P(y_j|x_i) - \sum_j P(y_j) \log P(y_j)$$

$$-H(y|x) + H(y) = H(y) - H(y|x)$$

$$I(x, y) = H(x) - H(x|y)$$

$$H(x) - H(x|y) = H(y) - H(y|x)$$

$$H(x,y) - H(y|x) - H(x|y)$$

$$H(x) + H(y) - H(x, y)$$

$$I(x_1, \cdots, x_N) = \sum_{i=1}^{N} H(x_i) - H(x_1, \cdots, x_N)$$

Q

$$H(P, Q) = \sum_i P_i \log_2 \frac{1}{Q_i} = -\sum_i P_i \log Q_i$$

$H(P)$

$H(P, Q)$

$$H(P) = -\sum_i P_i \log P_i \leq H(P,Q) = -\sum_i P_i \log Q_i$$

$$H(P, Q) = H(P)$$

$$Q = P$$

$$\log x \leq x - 1$$

$$\sum_i P_i \log\left(\frac{Q_i}{P_i}\right) \le \sum_i P_i\left(\frac{Q_i}{P_i} - 1\right) = \sum_i Q_i - \sum_i P_i = 0$$

$$\sum_i P_i \log\left(\frac{Q_i}{P_i}\right) = \sum_i P_i \log\left(\frac{1}{P_i}\right) + \sum_i P_i \log Q_i = H(P) - H(P,Q) \leq 0$$

$$KL(P||Q)$$

$$H(P,Q) - H(P) = -\sum_i P_i \log Q_i - (-\sum_i P_i \log P_i)$$

$$\sum_i P_i(\log P_i - \log Q_i) = \sum_i P_i \log\left(\frac{P_i}{Q_i}\right) \geq 0$$

$$KL(P||Q)$$

$$KL(P||Q) \neq KL(Q||P)$$

$$I(x,y) = \sum_i \sum_j P(x_i, y_j) \log \frac{P(x_i, y_j)}{P(x_i)P(y_j)} = KL(P(x,y)||P(x)\,P(y))$$

$$P(x)\,P(y)$$

$$P(x, y)$$

$$P(x, y) = P(x)\, P(y), \qquad P(x|y) = P(x), \qquad P(y|x) = P(y)$$

$g(x)$

$H(p, g)$

$$-\int p(x)\log g(x)\,dx = -\int p(x)\left(log\frac{1}{\sqrt{2\pi\sigma^2}} - \frac{(x-\mu)^2}{2\sigma^2}\right)dx$$

$$\frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\int (x-\mu)^2 p(x)\, dx = \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sigma^2$$

$$\frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2} = \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2}\log e = \frac{1}{2}\log(2\pi e\sigma^2) = H(g)$$

$H(p, g)$

$$H(p, g) \geq H(p)$$

$$H(g) = H(p, g) \geq H(p)$$

$$N(x, \mu, \sigma^2)$$

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

$$P(B) = \sum_{i=1}^{n} P(B|A_i)\, P(A_i), \qquad P(A_i|B) = \frac{P(B|A_i)\, P(A_i)}{P(B)} = \frac{P(B|A_i)\, P(A_i)}{\sum_{i=1}^{n} P(B|A_i)\, P(A_i)} \quad (i = 1, \cdots, n)$$

$$P(A_i)$$

A_i

$$P(B|A_i)$$

$$P(A_i|B)$$

D

$$= \frac{\text{likelihood} \times \text{prior}}{evidence}$$

$$p(H|D) = \frac{p(D|H)\,p(H)}{p(D)} \propto p(D|H)\,p(H)$$

$p(H)$

$$p(H|D)$$

$$p(D|H)$$

$p(D)$

$\theta_1, \cdots, \theta_M$

$$\theta = [\theta_1, \cdots, \theta_M]^T$$

$$\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N]$$

d

$$\mathbf{x}_n = [x_1, \cdots, x_d]^T$$

α

$$p(\theta|\mathbf{X}) = \frac{p(\mathbf{X}|\theta)\, p(\theta)}{p(\mathbf{X})} = \frac{p(\mathbf{X}|\theta)\, p(\theta)}{\sum_\theta p(\mathbf{X}|\theta)\, p(\theta)} \propto p(\mathbf{X}|\theta)\, p(\theta)$$

$p(\theta)$

θ

$$L(\theta|\mathbf{X}) = p(\mathbf{X}|\theta)$$

$$p(\mathbf{X}|\theta) = \prod_{n=1}^{N} p(\mathbf{x}_n|\theta)$$

$$p(\mathbf{X}) = \sum_{\theta} p(\mathbf{X}|\theta)p(\theta)$$

θ

$$p(\mathbf{X})$$

$$p(\mathbf{X}|\theta)\, p(\theta)$$

$$p(\theta|\mathbf{X})$$

$$\theta_{MLE} = \arg\max_{\theta} \; p(\mathbf{X}|\theta)$$

$$\theta_{MAP} = \arg\max_{\theta} \; p(\theta|\mathbf{X}) \propto p(\mathbf{X}|\theta)\, p(\theta)$$

$$p(\theta) = costant$$

$$\theta_{MLE} = \arg\max_{\theta} \log\ p(\mathbf{X}|\theta) = \arg\max_{\theta} \log\left(\prod_{n=1}^{N} p(\mathbf{x}_n|\theta)\right) = \arg\max_{\theta}\left(\sum_{n=1}^{N} \log\ p(\mathbf{x}_n|\theta)\right)$$

$$\theta_{MAP} = \arg\max_{\theta} \log \ (p(\mathbf{X}|\theta)\, p(\theta)) = \arg\max_{\theta} \left(\sum_{n=1}^{N} \log \ p(\mathbf{x}_n|\theta) + \log \ p(\theta) \right)$$

$$Q(\theta)$$

θ_n

θ_{n+1}

$$\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i),\ i = 1, \cdots, N\}$$

$$\mathbf{y} = \mathbf{f}(\mathbf{x}, \theta)$$

𝒟

$$L(\theta|\mathcal{D}) = p(\mathcal{D}|\theta) = \prod_{i=1}^{N} Pr(\mathbf{y} = \mathbf{y}_i|\mathbf{x} = \mathbf{x}_i; \theta)$$

$$l(\theta|\mathcal{D}) = \log L(\theta|\mathcal{D})$$

$$\log\left(\prod_{i=1}^{N} Pr(\mathbf{y}=\mathbf{y}_i|\mathbf{x}=\mathbf{x}_i;\theta)\right)$$

$$\sum_{i=1}^{N} \log Pr\left(\mathbf{y} = \mathbf{y}_i | \mathbf{x} = \mathbf{x}_i; \theta\right)$$

$$L(\theta|\mathcal{D})$$

$$l(\theta|\mathcal{D})$$

$\mathbf{y}_i$

$p(\mathbf{x}, \mathbf{y})$

$$N \longrightarrow \infty$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \log Pr(\mathbf{y} = \mathbf{y}_i | \mathbf{x} = \mathbf{x}_i; \theta) = \sum_i \sum_j p(\mathbf{x}_i, \mathbf{y}_j) \log p(\mathbf{y}_j | \mathbf{x}_i; \theta)$$

$$\sum_i \sum_j p(\mathbf{x}_i, \mathbf{y}_j) \log \left[\frac{p(\mathbf{y}_j|\mathbf{x}_i;\theta)}{p(\mathbf{y}_j|\mathbf{x}_i)} p(\mathbf{y}_j|\mathbf{x}_i) \right] = \sum_i \sum_j p(\mathbf{x}_i, \mathbf{y}_j) \left[-\log \frac{p(\mathbf{y}_j|\mathbf{x}_i)}{p(\mathbf{y}_j|\mathbf{x}_i;\theta)} + \log(\mathbf{y}_j|\mathbf{x}_i) \right]$$

$$-\sum_i \sum_j p(\mathbf{x}_i, \mathbf{y}_j) \log \frac{p(\mathbf{y}_j|\mathbf{x}_i)}{p(\mathbf{y}_j|\mathbf{x}_i;\theta)} + \sum_i \sum_j p(\mathbf{x}_i, \mathbf{y}_j) \log(\mathbf{y}_j|\mathbf{x}_i)$$

$$-KL(\mathbf{y}||\mathbf{y}_\theta) - H(\mathbf{y}|\mathbf{x})$$

$$H(\mathbf{y}|\mathbf{x})$$

$$KL(\mathbf{y}||\mathbf{y}_\theta)$$

$$p(\mathbf{y}|\mathbf{x};\theta)$$

$$p(\mathbf{y}_j|\mathbf{x}_i)$$

$$L(\theta|\mathcal{D})$$

$$\theta = \{\mathbf{m}, \mathbf{\Sigma}\}$$

$$\mathcal{N}(\mathbf{x}, \mathbf{m}, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{n/2} \left|\mathbf{\Sigma}\right|^{1/2}} \ \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{m})\right]$$

$$\mathbf{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$$

$$L(\theta|\mathbf{X}) = L(\mathbf{m}, \mathbf{\Sigma}|\mathbf{X}) = p(\mathbf{X}|\mathbf{m}, \mathbf{\Sigma}) = \prod_{i=1}^{N} \mathcal{N}(\mathbf{x}_i|\mathbf{m}, \mathbf{\Sigma})$$

$$l(\mathbf{m}, \mathbf{\Sigma}|\mathbf{X})$$

$$\log\ L(\mathbf{m},\boldsymbol{\Sigma}|\mathbf{X}) = \log\left[\prod_{i=1}^{N}\mathcal{N}(\mathbf{x}_i|\mathbf{m},\boldsymbol{\Sigma})\right] = \sum_{i=1}^{N}\log\mathcal{N}(\mathbf{x}_i|\mathbf{m},\boldsymbol{\Sigma})$$

$$\sum_{i=1}^{N}\left[-\frac{d}{2}\log 2\pi-\frac{1}{2}\log|\mathbf{\Sigma}|-\frac{1}{2}(\mathbf{x}_i-\mathbf{m})^T\mathbf{\Sigma}^{-1}(\mathbf{x}_i-\mathbf{m})\right]$$

$$\frac{\partial}{\partial \mathbf{m}} l(\mathbf{m}, \boldsymbol{\Sigma} | \mathbf{X})$$

$$-\frac{1}{2}\sum_{i=1}^{N}\frac{\partial}{\partial \mathbf{m}}\left((\mathbf{x}_i-\mathbf{m})^T\mathbf{\Sigma}^{-1}(\mathbf{x}_i-\mathbf{m})\right)$$

$$-\frac{1}{2}\sum_{i=1}^{N}\mathbf{\Sigma}^{-1}(\mathbf{x}_i-\mathbf{m})=\mathbf{0}$$

$$-2\mathbf{\Sigma}$$

$$\sum_{i=1}^{N} (\mathbf{x}_i - \mathbf{m}) = \mathbf{0}$$

$$\hat{\mathbf{m}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i$$

$$\frac{\partial}{\partial \boldsymbol{\Sigma}} l(\mathbf{m}, \boldsymbol{\Sigma} | \mathbf{X})$$

$$-\frac{N}{2}\frac{\partial}{\partial\boldsymbol{\Sigma}}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{N}\frac{\partial}{\partial\boldsymbol{\Sigma}}\left((\mathbf{x}_i-\mathbf{m})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\mathbf{m})\right)$$

$$-\frac{N}{2}\mathbf{\Sigma}^{-1} + \frac{1}{2}\sum_{i=1}^{N}\mathbf{\Sigma}^{-1}(\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T\mathbf{\Sigma}^{-1} = \mathbf{0}$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T$$

$$\frac{d}{d\mathbf{A}} \log |\mathbf{A}| = (\mathbf{A}^{-1})^T, \qquad \frac{d}{d\mathbf{A}} \left(\mathbf{a}^T \mathbf{A}^{-1} \mathbf{b} \right) = -(\mathbf{A}^{-1})^T \mathbf{a} \mathbf{b}^T (\mathbf{A}^{-1})^T$$

$$N - 1$$

$$d > N - 1$$

$$\mathbf{x} = [x_1, \cdots, x_N]^T$$

$$\mathbf{\Sigma}_x \phi_n = \lambda_n \phi_n \qquad (n = 1, \cdots, N)$$

$$\lambda_N \geq \cdots \geq \lambda_1 \geq 0, \qquad \phi_i^T \phi_j = \delta_{ij} = \begin{cases} 0 & i \geq j \\ 1 & i = j \end{cases} \qquad (i, j = 1, \cdots, N)$$

$$\mathbf{\Phi}^T = \mathbf{\Phi}^{-1}$$

$$\mathbf{\Sigma\Phi} = \mathbf{\Phi\Lambda},$$

$$\mathbf{\Phi}^{-1}\mathbf{\Sigma}\mathbf{\Phi} = \mathbf{\Phi}^T\mathbf{\Sigma}\mathbf{\Phi} = \mathbf{\Lambda}$$

$$\mathbf{\Lambda} = diag(\lambda_1, \cdots, \lambda_N), \qquad \mathbf{\Phi} = [\phi_1, \cdots, \phi_N]$$

Φ

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \mathbf{\Phi}^T \mathbf{x} = \begin{bmatrix} \phi_1^T \\ \vdots \\ \phi_N^T \end{bmatrix} \mathbf{x}$$

$$y_n = \phi_n^T \mathbf{x}$$

ϕ_n

$$\mathbf{x} = \mathbf{\Phi y} = [\phi_1, \cdots, \phi_N] \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \sum_{n=1}^{N} y_n \phi_n$$

$\phi_1, \cdots, \phi_N$

$$\mathbf{\Sigma}_y = \mathbf{\Phi}^T \mathbf{\Sigma}_x \mathbf{\Phi} = \mathbf{\Phi}^{-1} \mathbf{\Sigma}_x \mathbf{\Phi} = \mathbf{\Lambda}$$

$$\begin{bmatrix} \sigma_{11}^2 & \cdots & \sigma_{1N}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{N1}^2 & \cdots & \sigma_{NN}^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_N \end{bmatrix}$$

$$y_i$$

y_j

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

$$\sigma_{ij}^2 = \begin{cases} \lambda_i & i = j \\ 0 & i \neq j \end{cases}$$

$M < N$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \mathbf{A}^T \mathbf{x} = \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_N^T \end{bmatrix} \mathbf{x}$$

$$y_i = \mathbf{a}_i^T \mathbf{x}$$

$$\mathbf{\Sigma}_y = \mathbf{A}^T \mathbf{\Sigma}_x \mathbf{A}$$

$$tr\boldsymbol{\Sigma}_{\mathbf{y}} = \sum_{i=1}^{N} \sigma_{y_i}^2 = \sum_{i=1}^{N} \sigma_{x_i}^2 = tr\boldsymbol{\Sigma}_x$$

A

$$\varepsilon_M(\mathbf{A}) = \sum_{i=1}^{M} \sigma_{y_i}^2 = \sum_{i=1}^{M} \mathbf{a}_i^T \mathbf{\Sigma}_x \mathbf{a}_i$$

$$\varepsilon_M(\mathbf{A})$$

$$\left\{ \begin{array}{l} \text{maximize: } \varepsilon_M(\mathbf{A}) = \sum_{i=1}^{M} \mathbf{a}_i^T \mathbf{\Sigma}_x \mathbf{a}_i \\ \text{subject to: } \quad \mathbf{a}_j^T \mathbf{a}_j = 1 \quad (j = 1, \cdots, N) \end{array} \right.$$

$$\mathbf{a}_j^T \mathbf{a}_j = 1$$

$$\frac{\partial}{\partial \mathbf{a}_i}\left[S_M(\mathbf{A}) - \sum_{j=1}^{M} \lambda_j(\mathbf{a}_j^T\mathbf{a}_j - 1)\right] = \frac{\partial}{\partial \mathbf{a}_i}\left[\sum_{j=1}^{M}(\mathbf{a}_j^T\boldsymbol{\Sigma}_x\mathbf{a}_j - \lambda_j\mathbf{a}_j^T\mathbf{a}_j + \lambda_j)\right]$$

$$\frac{\partial}{\partial \mathbf{a}_i}\left(\mathbf{a}_i^T \mathbf{\Sigma}_x \mathbf{a}_i - \lambda_i \mathbf{a}_i^T \mathbf{a}_i\right) \overset{*}{=} 2\left(\mathbf{\Sigma}_x \mathbf{a}_i - \lambda_i \mathbf{a}_i\right) = 0, \qquad (i = 1, \cdots, N)$$

$\mathbf{a}_i$

ϕ_i

$$\mathbf{A} = \mathbf{\Phi}$$

$$\varepsilon_M(\mathbf{A}) = \varepsilon_M(\mathbf{\Phi}) = \sum_{i=1}^{M} \phi_i^T \mathbf{\Sigma}_x \phi_i = \sum_{i=1}^{M} \lambda_i$$

M

$$\lambda_1 \geq \cdots \geq \lambda_M \geq \cdots \geq \lambda_N$$

$$\mathbf{H}_f(\mathbf{x}) \geq 0$$

$$f_1(x) = x^2$$

$$f_2(x) = e^x$$

$$f_1''(x) = 2$$

$$f_2''(x) = e^x$$

$$c \in [0,\ 1]$$

$$f(c\mathbf{x}_1 + (1-c)\mathbf{x}_2) \leq c\, f(\mathbf{x}_1) + (1-c) f(\mathbf{x}_2)$$

$$c\mathbf{x}_1 + (1 - c)\mathbf{x}_2$$

$f(\mathbf{x}_1)$

$f(\mathbf{x}_2)$

$$g(x) = -f(\mathbf{x})$$

$$g(c\mathbf{x}_1 + (1-c)\mathbf{x}_2) \geq c\,g(\mathbf{x}_1) + (1-c)g(\mathbf{x}_2)$$

$$f\left(\sum_{i=1}^{n} \lambda_i \mathbf{x}_i\right) \leq \sum_{i=1}^{n} \lambda_i f(\mathbf{x}_i)$$

$$g(\mathbf{x})$$

$$g\left(\sum_{i=1}^{n} \lambda_i \mathbf{x}_i\right) \geq \sum_{i=1}^{n} \lambda_i g(\mathbf{x}_i)$$

λ_1

$$\lambda_2$$

$$\lambda_1 + \lambda_2 = 1$$

$$f(\lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2) \leq \lambda_1 f(\mathbf{x}_1) + \lambda_2 f(\mathbf{x}_2)$$

k

$$f\left(\sum_{i=1}^{k}\lambda_i\mathbf{x}_i\right) \leq \sum_{i=1}^{k}\lambda_i f(\mathbf{x}_i)$$

$k+1$

$$f\left(\sum_{i=1}^{k+1}\lambda_i\mathbf{x}_i\right) = f\left(\lambda_a\mathbf{x}_1 + (1-\lambda_1)\sum_{i=2}^{k+1}\frac{\lambda_i}{1-\lambda_1}\mathbf{x}_i\right) \leq \lambda_1 f(\mathbf{x}_1) + (1-\lambda_1) f\left(\sum_{i=2}^{k+1}\frac{\lambda_i}{1-\lambda_1}\mathbf{x}_i\right)$$

$$\sum_{i=2}^{k+1} \frac{\lambda_i}{1-\lambda_1} = 1$$

$$f\left(\sum_{i=2}^{k+1} \frac{\lambda_i}{1-\lambda_1} \mathbf{x}_i\right) \leq \sum_{i=2}^{k+1} \frac{\lambda_i}{1-\lambda_1} f(\mathbf{x}_i)$$

$$f\left(\sum_{i=1}^{k+1}\lambda_i\mathbf{x}_i\right) \leq \lambda_1 f(\mathbf{x}_1) + (1-\lambda_1)\left(\sum_{i=2}^{k+1}\frac{\lambda_i}{1-\lambda_1}f(\mathbf{x}_i)\right) = \sum_{i=1}^{k+1}\lambda_i f(\mathbf{x}_i)$$

$$f(E(\mathbf{x})) \leq E(f(\mathbf{x}))$$

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

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \quad \sim \quad \mathcal{N}(\mu, \sigma^2/N)$$

$\bar{x}$

$\boldsymbol{\mu}$

σ^2/N

$$N = 1$$

$\mathcal{z}$

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{N}} \quad \sim \quad \mathcal{N}(0, 1)$$

σ

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N} e_i^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N} x_i - N\bar{x}\right]$$

Z

T

$$t = \frac{\bar{x} - \mu}{S/\sqrt{N}} \quad \sim \quad T_\nu(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}$$

$$\Gamma(n) = (n-1)!$$

$$\nu = N - 1$$

$$T_1(t) = \frac{1}{\pi(1+t^2)}, \quad T_2(t) = \frac{1}{(2+t^2)^{3/2}}, \quad T_3(t) = \frac{6\sqrt{3}}{\pi(3+t^2)^2}, \quad T_\infty(t) = \frac{1}{\sqrt{2\pi}} e^{t^2/2}$$

$\nu \to \infty$

$$(1-\alpha)100\%$$

$$P(L \leq \mu \leq U) = (1 - \alpha)100\%$$

$$\alpha = 0.05$$

$$P(L \leq \mu \leq U) = 95\%$$

$$\left[-z_{\alpha/2},\ z_{\alpha/2}\right]$$

$$P(-z_{\alpha/2} \leq z \leq z_{\alpha/2}) = \int_{z_{-\alpha/2}}^{z_{\alpha/2}} \mathcal{N}(z, 0, 1)\, dz = 1 - \alpha,$$

$$\int_{-\infty}^{z_{-\alpha/2}} \mathcal{N}(z,0,1)\,dz = \int_{z_{\alpha/2}}^{\infty} \mathcal{N}(z,0,1)\,dz = \alpha/2$$

$$z_{\alpha/2} = 1.96$$

$$P(-z_{\alpha/2} \leq z \leq z_{\alpha/2})$$

$$P\left(-z_{\alpha/2} \leq \frac{\bar{x}-\mu}{\sigma/\sqrt{N}} \leq z_{\alpha/2}\right)$$

$$P\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}\right) = 1 - \alpha$$

$$1-\alpha$$

$$L = \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{N}}, \qquad U = \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{N}}$$

$$P(L \leq \mu \leq U) = 1 - \alpha$$

$\sigma/\sqrt{N}$

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{N}} = \bar{x} \pm z_{\alpha/2} SE$$

$$ESE = S/\sqrt{N}$$

t

$$\bar{x} \pm t_{\alpha/2} \frac{S}{\sqrt{N}} = \bar{x} \pm t_{\alpha/2} ESE$$

$t_{\alpha/2}$

$$P(-t_{\alpha/2} \leq t \leq t_{\alpha/2}) = \int_{t_{-\alpha/2}}^{t_{\alpha/2}} T_{\nu}(t)dt = 1 - \alpha$$

ν

$N = 11$

$$\bar{x} = \frac{1}{11} \sum_{i=1}^{11} x_i$$

$$\nu = N - 1 = 10$$

$$t_{0.05/2} = 2.228$$

$$t_{0.05/2} = 1.96$$

$$T_{\nu}(t) = T_{N-1}(t)$$

$$N(z, 0, 1)$$

$$t_{\alpha/2} \longrightarrow z_{\alpha/2}$$