$$\boxed{p(\theta \mid D) = \frac{p(\theta) \cdot p(D \mid \theta)}{p(D)}}$$

$| X$

$$D = \{X, y\}$$

$$\boxed{p(\theta \mid X, y) = \frac{p(\theta) \cdot p(y \mid X, \theta)}{p(y \mid X)}}$$

$$\frac{p(\theta) \cdot \boxed{p(y, x \mid \theta)}}{\underbrace{p(y \mid x)}_{= p(y \mid x) \cdot \cancel{p(x)}}}$$

$$\boxed{p(y \mid x, \theta)} \quad \cancel{p(x)}$$

$C = 4$ classes

$y = 1 \quad t = [1 \quad 0 \quad 0 \quad 0]$

$y = 2 \quad t = [0 \quad 1 \quad 0 \quad 0]$

$y = 3 \quad t = [0 \quad 0 \quad 1 \quad 0]$

$y = 4 \quad t = [0 \quad 0 \quad 0 \quad 1]$

$$p(C_1|x) = \frac{p(C_1, x)}{p(x)}$$

$$= \frac{p(x|C_1) \cdot p(C_1)}{p(C_1)\, p(x|C_1) + p(C_2)\, p(x|C_2)}$$

$$= \sigma(a) \qquad a = \ln \frac{p(x|C_1) \cdot p(C_1)}{p(x|C_2) \cdot p(C_2)}$$

$$p(x|C_1) = \mathcal{N}(x; \mu_1, \Sigma) \qquad p(C_1) = \pi_1$$

$$p(x|C_2) = \mathcal{N}(x; \mu_2, \Sigma) \qquad p(C_2) = 1 - \pi_1$$

$\boxed{a\,?}$

$$a = \ln p(x|C_1) - \ln p(x|C_2) + \ln p(C_1) - \ln p(C_2)$$

$$\ln p(x|C_1) = \cancel{-\frac{D}{2}\ln(2\pi)} - \cancel{\frac{1}{2}\log|\Sigma|}$$

$$-\frac{1}{2}(x-\mu_1)^T \Sigma^{-1}(x-\mu_1)$$

$$\ln p(x|C_2) = \cancel{-\frac{D}{2}\ln(2\pi)} - \cancel{\frac{1}{2}\log|\Sigma|}$$

$$-\frac{1}{2}(x-\mu_2)^T \Sigma^{-1}(x-\mu_2)$$

$$\ln p(x|C_1) - \ln p(x|C_2)$$

$$= x^T(\Sigma^{-1}\mu_1 - \Sigma^{-1}\mu_2)$$

$$-\frac{1}{2}\mu_1\Sigma^{-1}\mu_1 + \frac{1}{2}\mu_2\Sigma^{-1}\mu_2$$

$$= x^T \overbrace{\Sigma^{-1}(\mu_1-\mu_2)}^{w}$$

$$\underbrace{-\frac{1}{2}\mu_1\Sigma^{-1}\mu_1 + \frac{1}{2}\mu_2\Sigma^{-1}\mu_2}_{b}$$

$$\boxed{a = x^T w + b + \ln(\pi_1) - \ln(1-\pi_1)}$$

$$\boxed{\underline{a} = \underbrace{x}_{} \underbrace{w}_{} + \underbrace{d}_{}}$$

$$f(x) = \theta^T x = x^T \underbrace{\theta}_{}$$

$$P_\theta(D) = \prod_n P_\theta(x_n, c_n)$$

$$= \prod_n P_\theta(x_n \mid c_n) \cdot P(c_n)$$

$$\theta = \operatorname{argmax} \underbrace{\log P_\theta(D)}_{}$$

$$p_\theta(x_n \mid c_n) = \mathcal{N}(x_n ; \mu_1, \Sigma)^{c_n}$$

$$\mathcal{N}(x_n ; \mu_2 \Sigma)^{\underline{1-c_n}}$$

$$p_\theta(c_n) = \pi_1^{c_n} \cdot (1-\pi_1)^{1-c_n}$$

$$\frac{dL}{d\mu_1}, \frac{dL}{d\mu_2}, \frac{dL}{d\Sigma}, \frac{dL}{d\pi_1} = 0$$

$$\boxed{\mu_1, \mu_2, \Sigma, \pi_1}$$

$$L = \ldots + \log|\Sigma| \qquad \Sigma^{-1}$$

$$\frac{d \log|\Sigma|}{d\Sigma} \qquad \frac{d\,\mathrm{tr}(\Sigma^{-1} A)}{d\Sigma}$$

Linear reg

- ML θ_{ML}
- MAP θ_{MAP}
- Bayes $p(\theta | D)$

$\mathcal{N}(\theta; \mu, \Sigma)$

Logistic reg

- ML
- MAP

optimisation

- Bayes ??

$$\boxed{p(\theta | X, y)} = \frac{p(\theta) \cdot \prod_n p(y_n | x_n, \theta)}{p(y | X)}$$

↳ analytically intractable

$$p(y^* \mid D, x^*) = \int p(\theta \mid x, y)\, p(y^* \mid \theta, x^*)\, d\theta$$

approximate
- Laplace [W5]
- Variational inference
- Expectation propagation

⋮

assymptotically exact
- Markov Chain Monte Carlo sampling

Variational inference VI

$$p(\theta|y,x) \approx q(\theta)$$

$$KL(q\|p) = \int q(\theta) \log \frac{q(\theta)}{p(\theta|y,x)} d\theta$$

min KL

$$KL \geq 0$$

$KL = 0$ when $q(\theta) = p(\theta|y,x)$

$$KL(q\|p) = \int q(\theta) \log \frac{q(\theta)}{\frac{p(\theta)\, p(y|x,\theta)}{p(y|x)}} d\theta$$

$$= \underbrace{\int q(\theta) \log \frac{q(\theta)}{p(\theta)\, p(y|x,\theta)} d\theta}_{-F} + \log p(y|x)$$

$$\boxed{F = \int q(\theta) \log \frac{p(\theta)\, p(y|x,\theta)}{q(\theta)} d\theta}$$

$$\min_{q(\theta)} KL \equiv \max_{q(\theta)} F$$

F : variational lower bound

$$KL = -F + \log p(y|x) \geq 0$$

$$\underline{F \leq \log p(y|x)}$$

Variational GPs

- Bayesian neural networks
- Variational autoencoders
- diffusion models

$$F = \int q(\theta) \log \frac{p(\theta)\, p(y \mid x, \theta)}{q(\theta)} \, d\theta$$

$$= -\mathrm{KL}(q(\theta) \,\|\, p(\theta)) + \int q(\theta) \log p(y \mid x, \theta) \, d\theta$$

$$= \frac{1}{K} \sum \log p(y \mid x, \theta_k)$$

$$\theta_k \sim q(\theta)$$

$$p(y^* \mid D, x^*) \simeq \int q(\theta)\, p(y^* \mid \theta, x^*)\, d\theta$$