generative vs Discriminative

Gaussian Discriminant Analysis (GDA)

09/09/24

$P(y|x)$

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

$P(x|y)P(y) = \dfrac{P(x|y)P(y)}{\sum_{y=c} P(x|y)P(y)}$

$P(x)$

$y=1 \quad y=0$

$y \sim Ber(\Phi)$

$P(y=1)=\Phi$

$P(y=0)=(1-\Phi)$

$\mathcal{N}(\mu_1, \Sigma_1) \quad \mathcal{N}(\mu_c, \Sigma_c)$

$l(\Phi, \mu_c, \Sigma_c) = \log P(x,y)$

$P(y_c) = \pi_c$

$= \log P(y_c) + \log P(x|y)$

$P(y|x)$

$= \underbrace{\log \pi_c - \frac{1}{2}\log(|\Sigma|)}_{} - \frac{1}{2}(x-\mu_c)^T \Sigma_c^{-1} (x-\mu_c)$

$= \gamma_c - \frac{1}{2} x^T \Sigma_c^{-1} x + x^T \Sigma_c^{-1} \mu_c - \underbrace{\frac{1}{2} \mu_c^T \Sigma^{-1} \mu_c}_{}$

$\Sigma_c = \Sigma$

Quadratic discriminant Analysis

$= \gamma_c - \frac{1}{2} x^T \Sigma_c^{-1} x + x^T \Sigma_c^{-1} \mu_c$

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

$$P(x;\mu,\Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

μ = mean ve

Σ = covaria

$|\Sigma|$ = Deter

$$P(y_c|x) = \gamma_c + \underline{x^T \Sigma^{-1} \mu_c} + K$$

$$= \underline{\gamma_c + x^T \beta_c} + \underline{K}$$

linear to the value of x

$$K = -\frac{1}{2} x^T \Sigma^{-1} x.$$

$$\beta_c = \Sigma^{-1} \mu_c$$

Generative vs Discriminative — Gaussian Discriminant Analysis (GDA) — 09/09/24

$$P(y_c|x) = \frac{e^{\gamma_c + x^T\beta_c}}{\sum_{c'} e^{\gamma_{c'} + x^T\beta_{c'}}}$$

$$P(y=1|x) = \frac{e^{\gamma_1 + x^T\beta_1}}{e^{\gamma_0 + x^T\beta_0} + e^{\gamma_1 + x^T\beta_1}}$$

$$= \frac{1}{1 + e^{\gamma_0 - \gamma_1 + x^T\beta_0 - x^T\beta_1}}$$

$$= \frac{1}{1 + e^{(\gamma_0 - \gamma_1) + x^T(\beta_0 - \beta_1)}}$$

$$= \sigma(\quad)$$

$$\boxed{P(y=1|x) = \sigma(\underline{w^T(x - x_0)})}$$

$$\theta^T x > 0$$

where $w = \beta_1 - \beta_0$

$$x_0 = \frac{1}{2}(\mu_1 + \mu_0) - (\mu_1 - \mu_0)\frac{\log(\pi_1/\pi_0)}{(\mu_1 - \mu_0)^T \Sigma^{-1} (\mu_1 - \mu_0)}$$

$$w^T(x - x_0) > 0$$

$$w^T x > \underbrace{w^T x_0}_{c}$$

Spam Email Classification — discrete feature

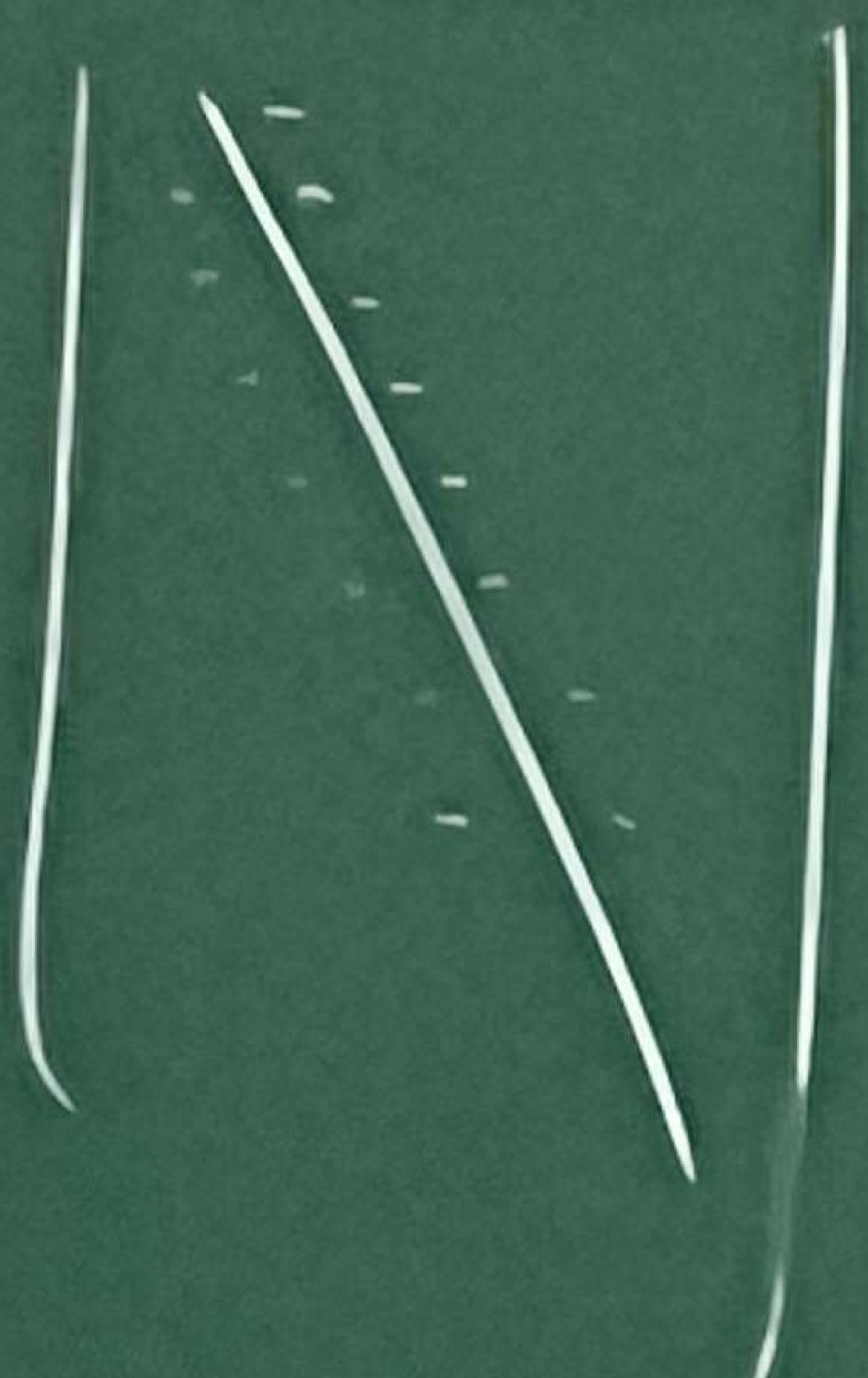
$$P(y|x) = \frac{P(x, y)}{P(x)} = \frac{P(\underline{x}|y)\,P(y)}{\cancel{P(x)}}$$

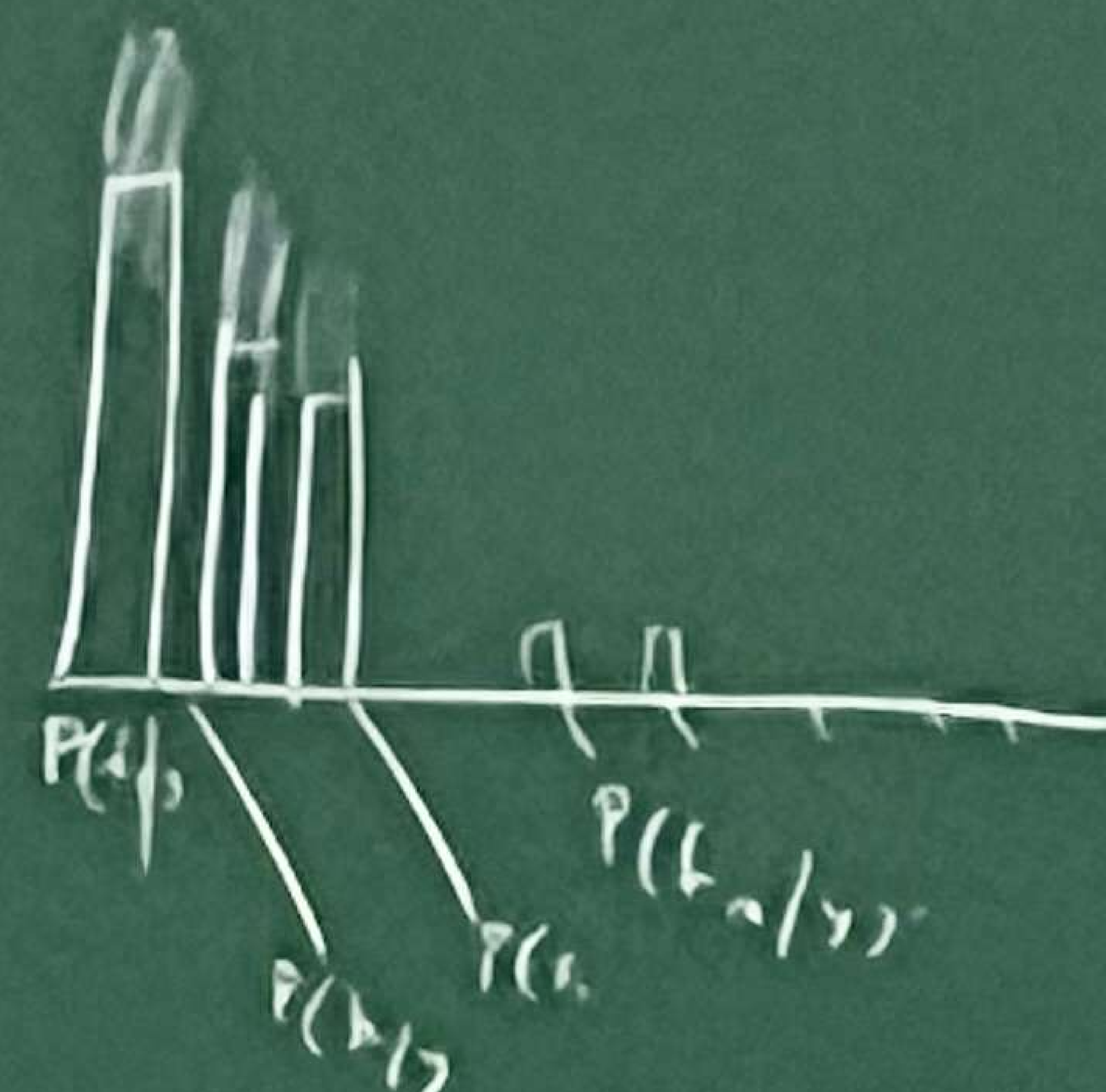
V = Vo

$$\frac{P(f_1|y)\,P(f_2|y) \ldots P(f_{10k}|y)\,P(y=1)}{P(x)} \qquad \frac{P(f_1, f_2, f_3 \ldots f_{10k})\,P(y)}{P(x)}$$

$$= P(f_1) \times P(f_2 | f_1) \ldots P(f_{10k} | f_1 f_2 \ldots f_{10k-1})$$

$\Sigma = \lambda I$ – NB





Smoothing

$$P(f_1/y) = 0$$

$$= \frac{\text{Count}(f_1, y) + 0.5}{\text{Count}(y) + 10K}$$

10K × 0.5

$$\sum_{i=1}^{10K} P(f_i/y) = 1$$