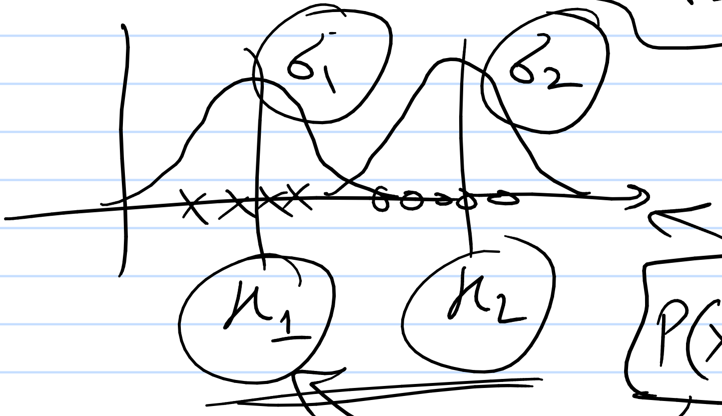
Gaussian model distribution, Gaussian mixture model (GMM)

$$1d \rightarrow P(X = x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\bar{\mu})^2}{2\sigma^2}\right)$$



$$P(G = g) = \begin{cases} 1 & p \\ 2 & 1-p \end{cases}$$

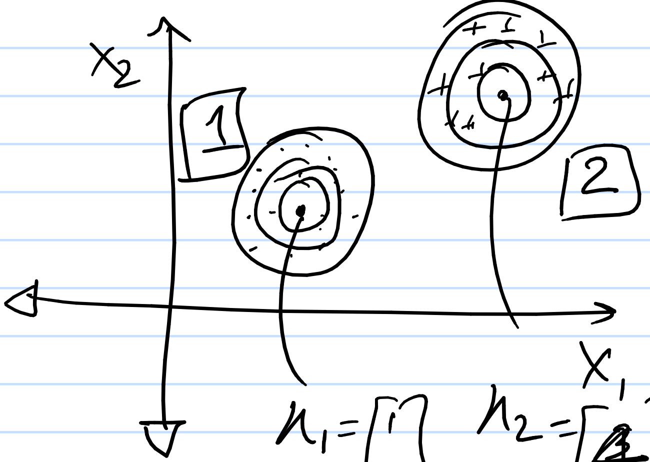
$P(X \mid G = g)$

$$\boxed{P(X = x \mid G = g)}$$

$$P(X = x \mid G = 1) = N(\mu_1, \sigma_1^2)$$

$$P(X = x \mid G = 2) = N(\mu_2, \sigma_2^2)$$

$$P(G = 1) = p$$



$\mu_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $\mu_2 = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$

$$p\left(X = \underbrace{\begin{bmatrix} x^1 \\ x^2 \end{bmatrix}}_{\underline{x}\ \text{2dim}} \middle| G = 1\right) = \frac{1}{\sqrt{2\pi}\,(\det \Sigma)^{1/2}} \exp\left(-\frac{(x-\mu)^T \Sigma^{-1} (x-\mu)}{2}\right)$$

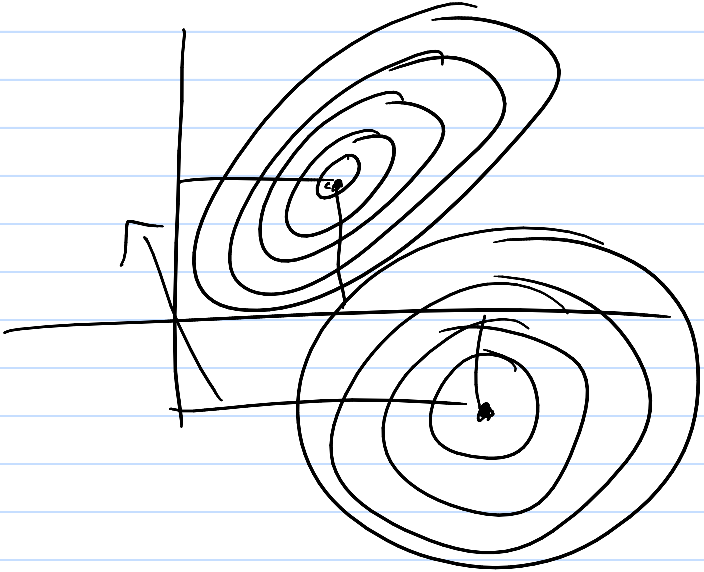
$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \sigma^2 = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$$

$$\boxed{\underbrace{(x-\mu)^T}_{1\times 2}\ \underbrace{\Sigma^{-1}}_{2\times 2}\ \underbrace{(x-\mu)}_{2\times 1}}\ 1\times 1$$

$$\begin{bmatrix} a & d \\ c & b \end{bmatrix}$$

Cov(W, 2)

$$\underbrace{\mathrm{Cov}(W, Z)} = \underline{E[(W - E[W])(Z - E[Z])]}$$



Mixture of 2 Gaussians.

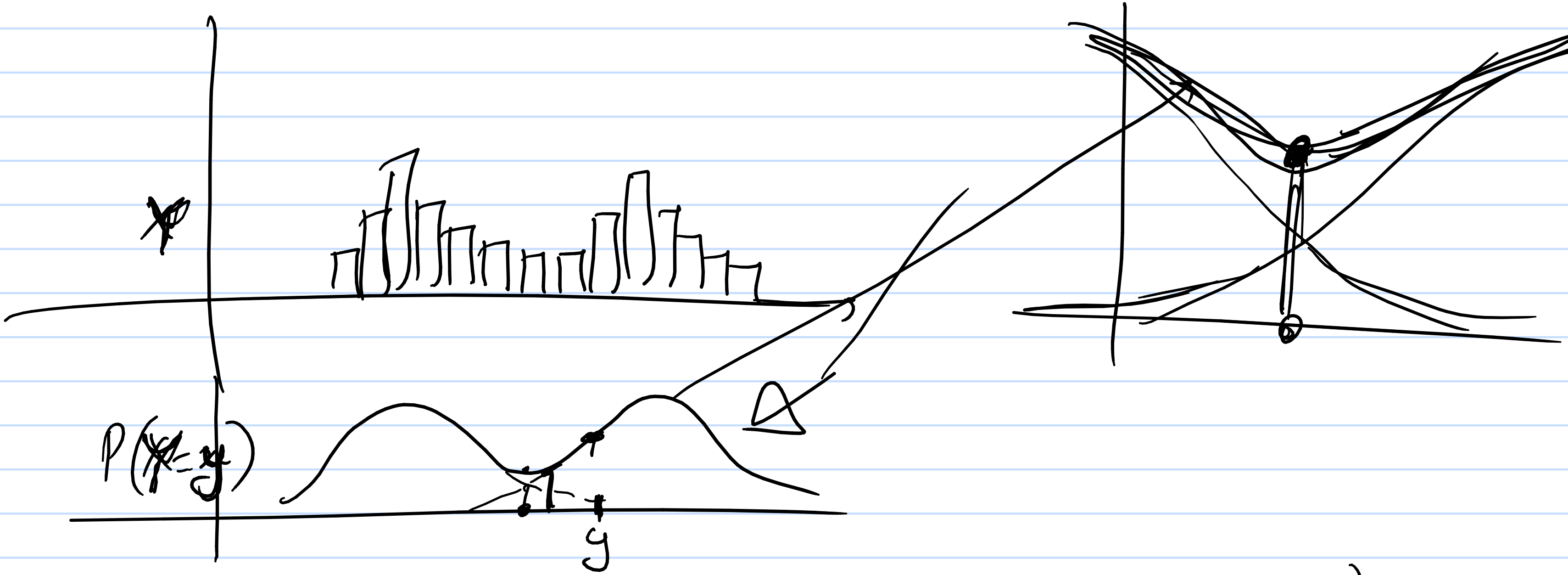
<u>Generative</u> description:

$\underline{\Delta} \sim \text{Bernoulli}(\pi)$

$Y_1 \sim N(\mu_1, \sigma_1^2)$

$Y_2 \sim N(\mu_2, \sigma_2^2)$

$Y = \underline{(1-\Delta)\, Y_1} + \underline{\Delta\, Y_2}$



$$\underline{P(Y=y)} = (1-\pi)\frac{1}{\sqrt{2\pi}\sigma_1} e^{\left(-\frac{(y-\mu_1)^2}{2\sigma_1^2}\right)} + \pi\cdot\frac{1}{\sqrt{2\pi}\sigma_2} e^{\left(-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right)}$$

$$\text{MLE}: \quad \sum_{i=1}^{N} \log P_\theta(Y = y_i)$$

$y_1 \ldots y_{100}$

$$\text{MLE} \equiv \min_{\mu} \sum_{i=1}^{N} (y_i - \mu)^2$$

$\log_e e^a = a$

$\log(e^a + e^b)$

EM

$P(Y=y)$

y_i, δ_i

$$P(Y=y, \Delta=\delta) = P_{\mu_1,\sigma_1}(Y=y)^{1-\delta} \cdot P_{\mu_2,\sigma_2}(Y=y)^{\delta}$$

0 or 1

$$\sum_i \log P(Y=y_i, \Delta=\delta_i) = \sum \boxed{(1-\delta_i)\log P_{\mu_1\sigma_1}(y_i) + \delta_i \cdot \log P_{\mu_2\sigma_2}(y_i)}$$

$(y_1, \dots \; y_N)$: data.

$\theta = (\mu_1, \sigma_1, \mu_2, \sigma_2, \pi)$

$\theta^0 \longrightarrow P(Y=y, \Delta=\delta)$

$P(\Delta=\delta_i \mid Y=y_i)$

Likelihood(θ) $\xrightarrow{\text{maxi}}$ $\mu_1', \mu_2', \sigma_1', \sigma_2', \pi'$

θ'

EM: θ^0

E: $P(Z^m | Z; \theta^0)$

M. $\max_{\theta} \sum_{Z^m} P(Z^m | Z, \theta^0) \cdot \ell(\theta; Z, Z^m)$

θ'

E: $P(Z^m | Z, \theta')$

M: $\max_{\theta}$

$\boxed{\max_{\theta} P(Z; \theta)}$

$$\log P(Z;\theta') = \log P(Z, Z^m; \theta') - \log P(Z^m \mid Z; \theta')$$

$$P(Z) = \frac{P(Z, Z^m)}{P(Z^m \mid Z)}$$

Take $E_{P(Z^m \mid Z;\theta)}$

$$L(\theta') = \sum_{Z^m} P(Z^m \mid Z;\theta) \cdot \log P(Z, Z^m; \theta')$$

$$- \sum_{Z^m} P(Z^m \mid Z;\theta) \cdot \log P(Z^m \mid Z;\theta')$$

$$Q(\theta', \theta) - R(\theta', \theta)$$

θ

$\downarrow$

θ' ~~is~~ maximizer.

in M Step.

$$L(\theta') - L(\theta) > 0$$

$$= Q(\theta', \theta) - R(\theta', \theta)$$

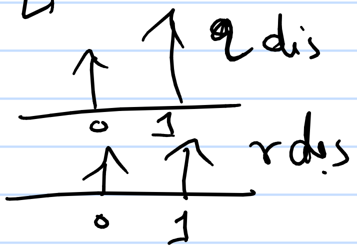
$$- \left(Q(\theta, \theta) - R(\theta, \theta) \right) > 0$$

$$= \underbrace{Q(\theta', \theta) - Q(\theta, \theta)}_{\geq 0} - \underbrace{\left(R(\theta', \theta) - R(\theta, \theta) \right)}_{> 0}$$

$$-\underline{\left(R(\theta',\theta) - R(\theta,\theta)\right)} > 0$$ irrespective of what θ'

$$R(\theta',\theta) = \sum_i q_i \log r_i = \sum_{\underbrace{z^m}_{i}} \underbrace{P(z^m|z;\theta)}_{q_i} . \log \underbrace{P(z^m|z;\theta')}_{r_i}$$

$$R(\theta,\theta) = \sum_i q_i \log q_i$$



$$\Rightarrow \quad \underline{\underline{-\sum_i q_i \log\left(\frac{r_i}{q_i}\right) > 0}}$$

$$\Leftrightarrow \quad \rightarrow \quad \sum_i q_i \log \frac{r_i}{q_i} < 0$$ irrespective of <u>any</u> r_i

$$\boxed{\max_{r_i} \sum_i q_i \log \frac{r_i}{q_i} \quad \text{subject to } \sum r_i = 1} \leq 0$$

$$r_i = q_i$$

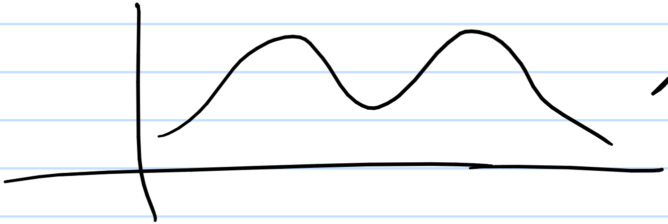
$$\frac{d}{dr_i}\left(\sum_i q_i \log \frac{r_i}{q_i} - \lambda\left(\sum r_i - 1\right)\right) = \frac{q_i}{r_i} - \lambda = 0$$

$$\frac{d}{d\lambda} \quad = 0 \Rightarrow \sum r_i = 1$$

$$q_i = \lambda r_i$$

$$1 = \sum q_i = \lambda \sum r_i \Rightarrow \lambda = 1$$

Gibbs Sampling



$\mu_1, \mu_2, \sigma_1, \sigma_2, \pi$

Bayes rule

Ridge regression

$\beta \sim N(0, \tau^2 I)$

$Y|\beta \sim N(\beta^T X, \sigma^2 I)$

$$P(\beta|Y) = \frac{P(Y|\beta)\, P(\beta)}{P(Y)}$$

$\int_\beta P(Y=y, \beta= \;)$

MAP estimate : most likely "β" value averaged prediction $= \boxed{\int_\beta \beta^T x^{test} \cdot P(\beta \mid \underbrace{data}_{Y}) \approx \frac{1}{100} \sum_j \beta_j^T x^{test}}$ $\beta_1 \ldots \beta_{100}$

Monte Carlo avg

Markov Chain Monte Carlo

$P(U_1, \ldots, U_p) \sim u_1, \ldots u_p$

Gibbs sampling

$[P(U_i \mid U_1 \ldots U_{i-1}, U_{i+1}, \ldots U_p)$ for every i

$U_1 \quad U_2 \quad U_3 \quad U_4$

$P_\theta(U_1, U_2, U_3, U_4)$ - target ★

$0 \quad 1 \quad -1 \quad 2$ ← 0th sample / observation

$\boxed{3} \quad \boxed{-5} \quad \boxed{0} \quad \boxed{}$ 1st sample. $\sim P(U_1 \mid U_2 = 1, U_3 = -1, U_4 = 2)$ ★

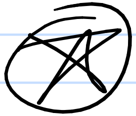
$\sim P(U_2 \mid U_1 = 3, U_3 = -1, U_4 = 2)$ ★

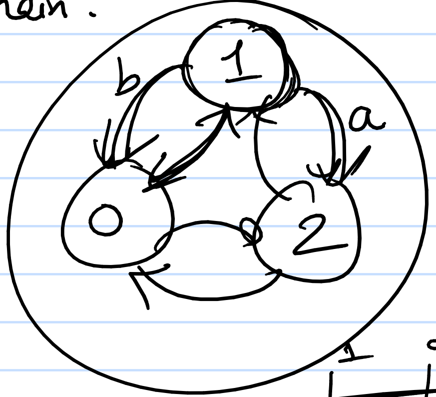
$\times$

$\times \quad \times \quad \boxed{5}$

Markov property : eg: $P(\underline{Z_{100}} \mid \underline{Z_{99}, Z_{98}, Z_{97} \ldots Z_1}) = P(Z_{100} \mid \underline{Z_{99}, Z_{98}, Z_{97}})$

Markov Chain :





$\underline{X_1 = 1}$

$P(X_2 = 0 \mid X_1 = 1)$

$P(X_2 = 2 \mid X_1 = 1)$

histogram

Stationary distribution

P

Transition distribution matrix

Q

$$\begin{array}{c|ccc} & 1 & 0 & 2 \\ \hline 1 \rightarrow & 0 & b & a \\ & & & \\ & & & \end{array}$$

3×3

GAM

(GLM)

$$E[Y|X] = c + f_1(X_1) \pm f_2(X_2) \pm \cdots \pm f_p(X_p)$$

$$\boxed{f_1(x_1) = a_1 x_1 + a_2 x_1^2 + a_3 x_1^3}$$

$$\mu = c + \sum_i f_i(X_i)$$

$$\log \mu,\ e^{\mu} = c + \sum_i f_i(X_i)$$

Log regression (binary)

$$\log\left(\frac{\boxed{P(G=1 \mid X=x)}}{1-P(G=1 \mid X=x)}\right) = \beta^T x$$

$$\boxed{E[G \mid X=x] = \mu}$$

$$\boxed{\log\left(\frac{\mu}{1-\mu}\right) = \beta^T x}$$ logit

$$\Phi^{-1}(\mu) = \beta^T x$$: probit