Density Estimation

Nearest neighbor estimator

$D = \{x_1, x_2, \ldots, x_N\} \in \mathbb{R}^d$

give $x \in \mathbb{R}^d$, estimate $\hat{P}(x)$ as follows:

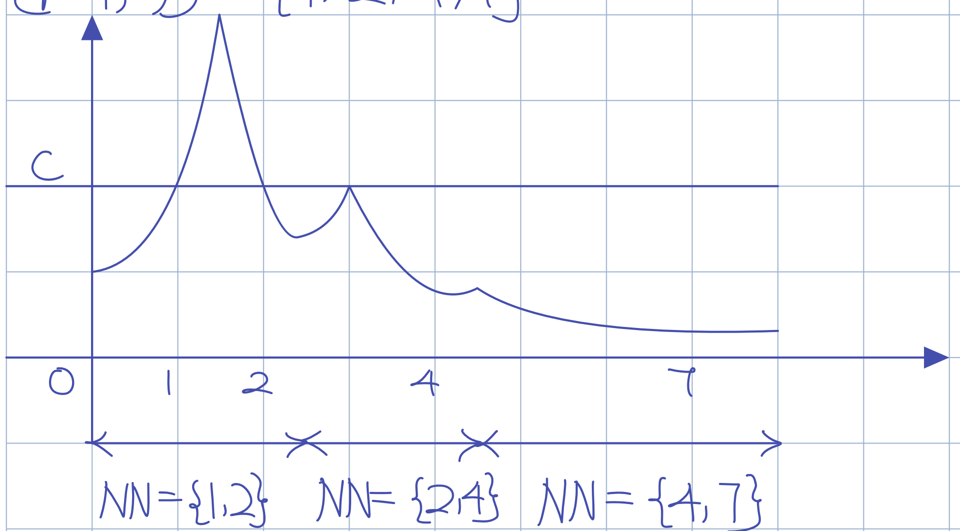
Let $x_{[1]}, x_{[2]}, \ldots, x_{[k]}$ be the k nearest neighbors of x

$d_k(x) = \| x - x_{[k]} \|$

$$\hat{P}(x) = \frac{C}{V_k(x)}$$

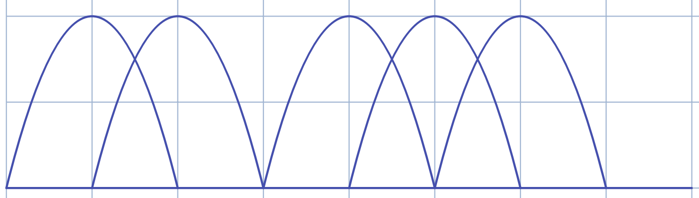
$V_k(x) \equiv$ Volume of sphere in $\mathbb{R}^d$ with radius $d_k(x)$

$d=1$, $D = \{1, 2, 4, 7\}$



$k=2$

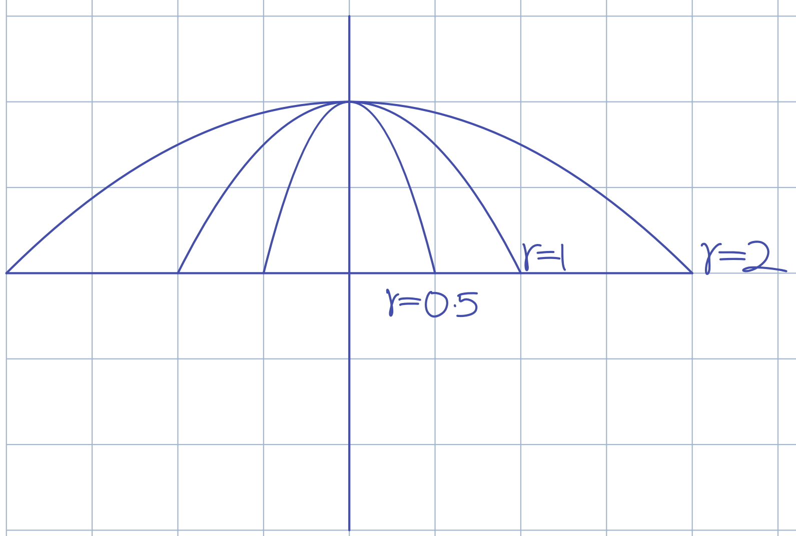
Parzen Window Estimation



Gaussian Kernel

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

$$\phi(z/r)$$



r = kernel weights

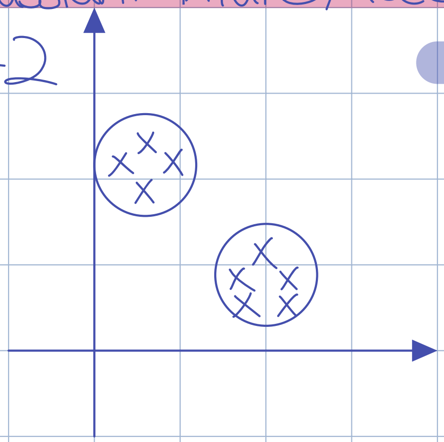
Give $D = \{x_1, \dots, x_N\}$

$$\hat{p}(x) = \frac{1}{k} \sum_{i=1}^{N} \phi\left(\frac{\|x - x_i\|}{r}\right)$$

r can be selected via validation

Gaussian Mixture Models

$d=2$



Multivariant Gaussian Distribution

Standard Normal:

$$Z \sim N(0,1)$$

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

$$Z \sim N(\mu, \sigma^2)$$

$$f_Z(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(z-\mu)^2/2\sigma^2}$$

μ = mean

σ = variance

Bivariant Gaussian Distribution

$$X_1 \sim N(\mu_1, \sigma_1^2)$$

$$X_2 \sim N(\mu_2, \sigma_2^2)$$

Covariant $E[(X_1-\mu_1)(X_2-\mu_2)] = \sigma_{12}$

Auto Correlation Coefficient $\rho = \dfrac{\sigma_{1,2}}{\sigma_1 \sigma_2}$ $\quad -1 \le \rho \le 1$

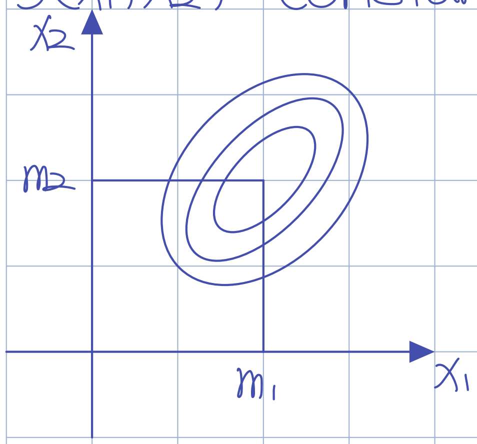
$\rho = 0$, zero correlation

Joint Density Function

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ \left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - \right.$$

$$2\rho \left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) \Big/ -2(1-\rho^2)$$

$f(x_1, x_2) = \text{constant}$



$$\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) = \text{constant}$$

Multivariant Gaussian

$$\underline{X} = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}$$

each x_i is Gaussian with mean μ_i and variance σ_i^2

Covariance Matrix

$$\underline{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d} \\ \vdots & & & \\ \sigma_{d1} & \sigma_{d2} & \cdots & \sigma_{dd} \end{bmatrix}$$

$$\sigma_{i,j} = E[(x_i-\mu_i)(x_j-\mu_j)]$$

$$\underline{\Sigma} = E[(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})^T]$$

mean vector

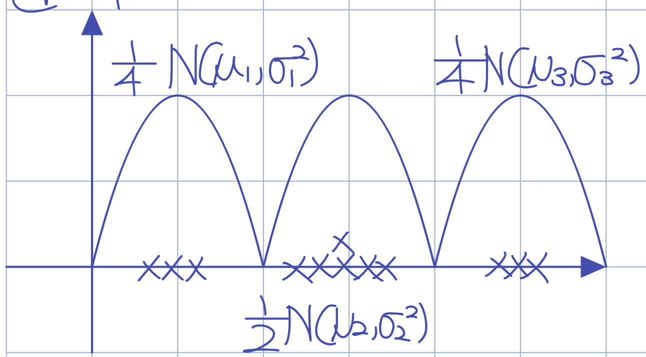
$$\underline{\mu} = \begin{bmatrix} m_1 \\ \vdots \\ m_d \end{bmatrix} \qquad f(\underline{x}) = \frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma|}} \exp\left\{\frac{-(x-\mu)^T \Sigma^{-1}(x-\mu)}{2}\right\}$$

det

Covariant Matrix

Gaussian Mixture Model

$d = 1$



Gaussian Mixture Model

$$\hat{p}(x) = \frac{1}{4} \times \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-(x-\mu_1)^2/2\sigma_1^2} + \frac{1}{2} \times \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-(x-\mu_2)^2/2\sigma_2^2} + \frac{1}{4} \times \frac{1}{\sqrt{2\pi\sigma_3^2}} e^{-(x-\mu_3)^2/2\sigma_3^2}$$

$$\hat{p}(x) = \frac{1}{4} N(x, \mu_1, \sigma_1^2) + \ldots$$

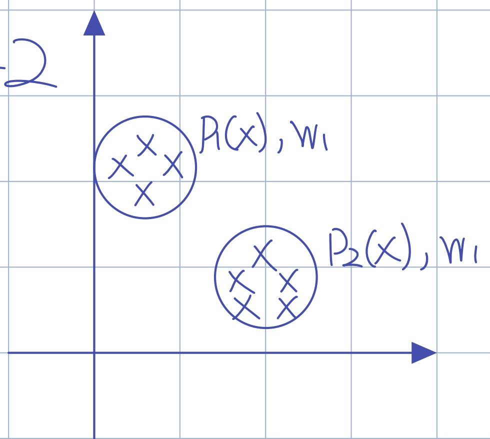
General Cases

$N(X, \mu_j, \Sigma_j)$ ← jth Gaussian Density Function $= P_j(x)$

$\hat{P}(x) = w_1 P_1(x) + \dots + w_K P_K(x)$

w_K = weight associated with $P_K(x)$

$d=2$



Gaussian Mixture Model

K = # of Gaussian

$P_j(x) = N(X, \mu_j, \Sigma_j)$

$$\hat{P}(x) = \sum_{j=1}^{K} \omega_j P_j(x)$$

Given $D = \{x_1, \dots, x_N\}$

$x_i \in \mathbb{R}^d$

K = # of Gaussians

Compute $\{\omega_j, \mu_j, \Sigma_j\}_{j=1}^{K} = \Omega$

Such that $\hat{P}_\Omega$

$$\hat{P}_\Omega(\underline{x}) = \sum_{j=1}^{K} \omega_j P_j(\underline{x})$$ is best fit for D

Find Ω that maximizes

$$\hat{P}_\Omega(\underline{x_1, \dots, x_N}) = \prod_{n=1}^{N} \hat{P}_\Omega(\underline{x_n})$$

Maximum likelihood solution

Equivalently

$$\log \hat{P}_\Omega(\underline{x_1, \dots, x_N}) = \log \prod_{n=1}^{N} \hat{P}_\Omega(\underline{x_n})$$

$$= \sum_{n=1}^{N} \log \hat{P}_\Omega(\underline{x_n})$$

Minimize

$$-\sum_{n=1}^{N} \log \hat{P}_\Omega(\underline{x_n})$$

Finding Ω that minimizes

$$E_{in}(\Omega) = \frac{1}{N} \sum_{n=1}^{N} \underbrace{-\log \hat{P}_\Omega(\underline{x_n})}_{e_n(\Omega)}$$

$$e_n(\Omega) = -\log\left(\sum_{j=1}^{K} \omega_j N(\underline{x_n}, \underline{\mu_j}, \Sigma_j)\right)$$

$$\Omega_{t+1} = \Omega_t - \varepsilon_t \nabla_{\Omega_t} e_n(\Omega_t)$$

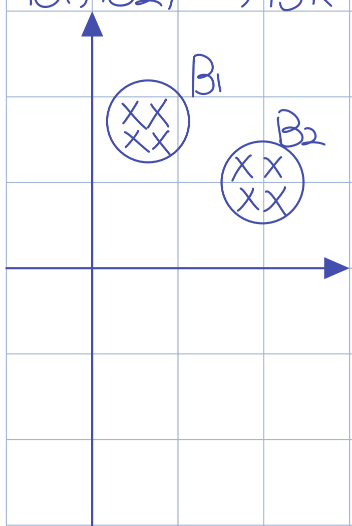
$$\omega_j \geq 0 \qquad \sum_{j=1}^{k} \omega_j = 1$$

Σ_j = covariant matrix

Expectation maximization algorithm (EM Algorithm)

Auxiliary Variables

$B_1, B_2, \dots, B_k \subseteq D$



B_i is the set of points in D sampled from $P_i(\underline{x})$

$\{B_i\}$ is disjoint

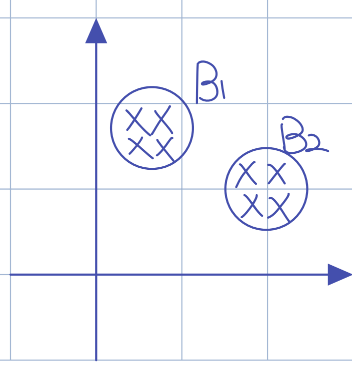
$B_1 \cup B_2 \cup \dots \cup B_k = D$

Subproblem #1

Suppose $B_1, B_2, \dots, B_k$ are given, find $\Omega = \{\omega_j, \mu_j, \Sigma_j\}_{j=1,2,\dots,k}$

$P_1(\underline{x}) = N(x_1, \mu_1, \Sigma_1), w_1$

$P_2(\underline{x}) = N(x_2, \mu_2, \Sigma_2), w_2$



$N_1 = \#$ Points In B_1

$N_2 = \#$ Points In B_2

Then $\omega_i = \dfrac{N_i}{N}$

$$\mu_i = \frac{1}{N_i} \sum_{x_n \in B_i} x_n$$

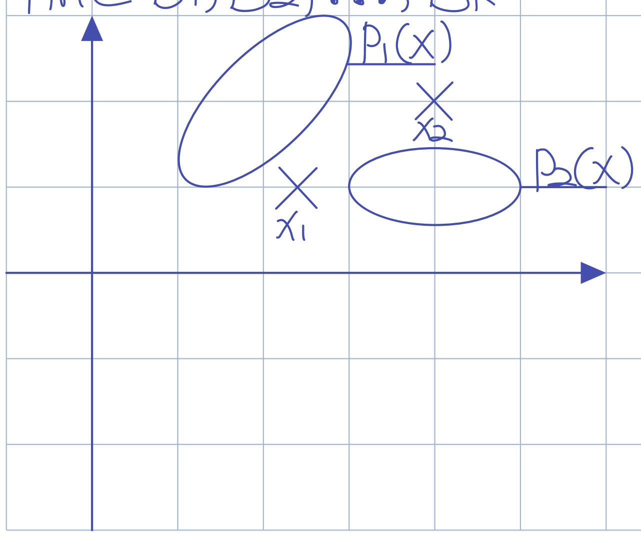
$$\Sigma_i = E\left[(x-\mu_i)(x-\mu_i)^T\right]$$

$$\Sigma_i = \frac{1}{N_i} \sum_{x_n \in B_i} (x_n - \mu_i)(x_n - \mu_i)^T$$

Subproblem 2

$$\Omega = \{\omega_j, \mu_j, \Sigma_j\}_{j=1,\dots,k}$$

Find $B_1, B_2, \dots, B_k$



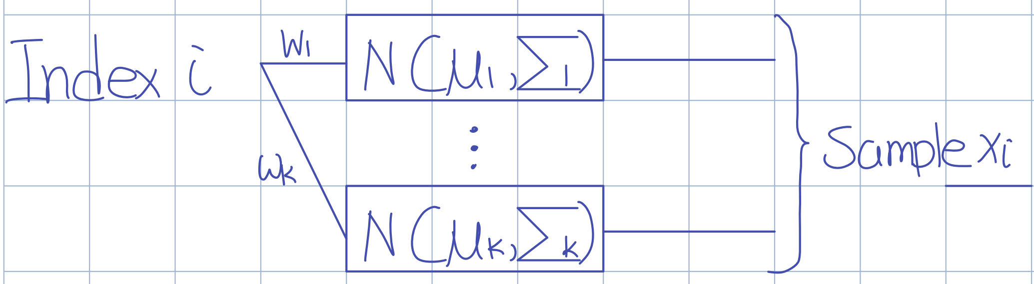
For each $x_i \in D$, is it sampled from $P_1(x)$ or $P_2(x)$ or $P_k(x)$

$$P(x_i \mid x_i \in B_j) = P_j(x_i)$$

$$P_r(\text{cluster} = B_j) = \omega_j$$

Question,

$$\Pr(\text{Cluster} = B_j \mid x_i) = ?$$



$$\Pr(\text{Cluster} = B_j \mid x_i) = \frac{P(x_i \mid B_j)\Pr(\text{Cluster} = B_j)}{\sum_{j=1}^{k} P(x_i \mid B_j)\Pr(B_j)}$$

$$= \frac{P_j(x_i)\,\omega_j}{\sum_{j=1}^{k} P_j(x_i)\,\omega_j}$$

Assign x_i to B_{j^*} where

$$j^* = \underset{1 \le j \le k}{\operatorname{argmax}} \Pr(\text{cluster} = B_j \mid x_i)$$

$$j^* = \underset{1 \le j \le k}{\operatorname{argmax}} P_j(x_i)\,\omega_j$$

$$\downarrow$$

$$\underline{N(x_i, \mu_j, \Sigma_j)}\,\omega_j$$

Recap

2 Subproblems: Given $B_1, \ldots, B_k$ Compute Ω

Given Ω, Compute $B_1, \ldots, B_k$

EM Algorithm, Alternate between subproblem 1 and 2 until convergence