Makine Öğrenmesi

Parametrik kestirim, MLE, Gaussian Mixtures, EM

Orijinal not: Ders-4.pdf

UYARI Bu belge, el yazısı ders notlarından Claude (Anthropic) yardımıyla dijitale aktarılmıştır. Yardımcı kaynak olarak hazırlanmıştır; hata veya eksik içerebilir. Lütfen orijinal notları ve ders kaynaklarını esas alın.

NOTICE *This document was transcribed from handwritten lecture notes with the help of Claude (Anthropic). It is meant as a supplementary resource and may contain errors or omissions. Always check against the original notes and official course material.*

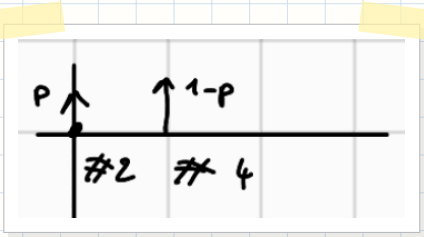
orj. s. 1

feature vektörü

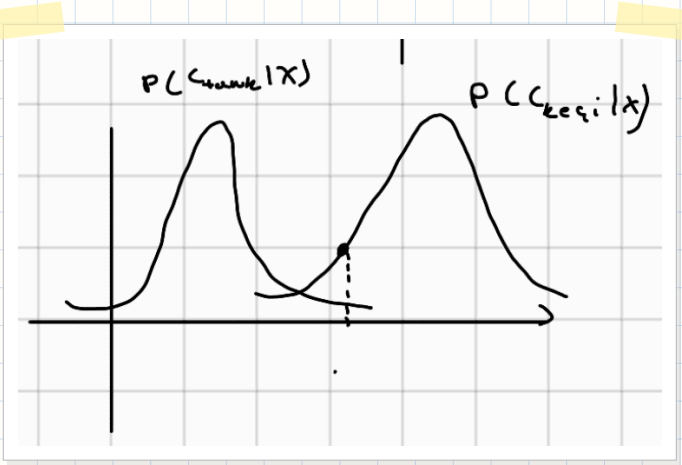
$$\vec{x} = \begin{bmatrix} \#\text{ayak sayısı} \\ \#\text{ağırlığı} \end{bmatrix} \qquad P(C_i|x)$$

$$\left.\begin{matrix} \text{tavuk} \\ \text{keçi} \end{matrix}\right\} \#\text{ayak sayısı} \qquad r^t \in \{2,4\}$$

Bernolli dağılımı:



orijinal çizim, s. 1



orijinal çizim, s. 1

Parametric kestirim:

① Varsayılan Model: θ

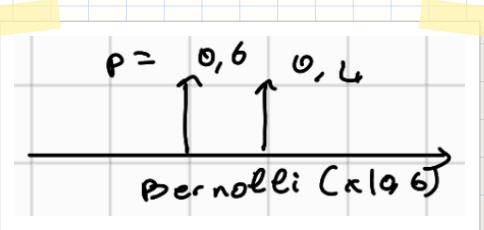
② En olabilirlik fonksiyonu: $L(\theta)$
③ Optimization Yöntemi: MLE maximum likelihood estimation

MLE:

$$\mathcal{X} = \{x^1, x^2, \dots, x^N\}$$

her bir örnek.

$$\underbrace{x^t}_{\text{bu dağılıma göre üretilir}} \sim p(x|\theta)$$



orijinal çizim, s. 1

$$L(\theta|\mathcal{X}) = p(x^1|\theta) \cdot p(x^2|\theta) \cdots$$
$$= \prod_{i=1}^{N} \underbrace{p(x^i|\theta)}_{0'\text{dan büyük}}$$

$\hat{\theta} = \arg\max_\theta L(\theta|\mathcal{X})$ = likelihood fonksiyonunu max yapan θ benim kestirim

orj. s. 2

logaritmasını alabiliriz

$$\ell(\theta|\mathcal{X}) = \log L(\theta|\mathcal{X})$$
$$= \sum_{i=1}^{N} \log p(x^i|\theta)$$

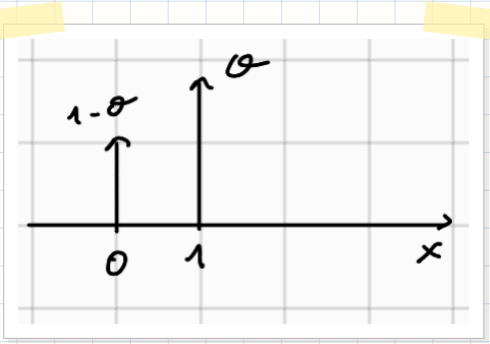
Yeterli istatistik (Sufficient Statistics)

Bir dağılımı açıklayabilecek minimum parametre sayısı

Bernoulli Distrubution:

$$P(x|\theta) = \theta^x (1-\theta)^{1-x}, \quad \underbrace{x \in \{0,1\}}_{\text{ya 0 ya 1}}, \quad \underbrace{\theta \in [0,1]}_{\text{0 ile 1 arasında}}$$

??



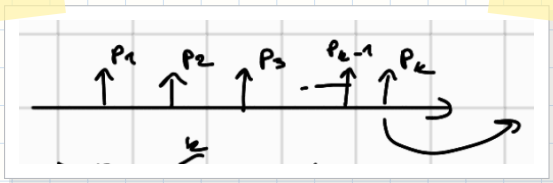
orijinal çizim, s. 2

$$L(\theta|\mathcal{X}) = \prod_{i=1}^{N} p(x^i|\theta)$$
$$= \prod_{i=1}^{N} \theta^{x_i}(1-\theta)^{(1-x_i)}$$
$$\ell(\theta|\mathcal{X}) = \sum_{i=1}^{N} \left[(x_i \log\theta) + (1-x_i)\log(1-\theta)\right]$$

$$x_i \in \{0,1\} \downarrow \quad \boxed{s = \sum_{i=1}^{N} x_i}$$

$$\frac{\partial \ell(\theta|\mathcal{X})}{\partial\theta} = 0 = \frac{\partial(s\log\theta)}{\partial\theta} + \frac{\partial\left(\log(1-\theta)(N-s)\right)}{\partial\theta}$$
$$= \frac{s}{\theta} - \frac{N-s}{1-\theta}$$
$$\frac{N-s}{1-\theta} = \frac{s}{\theta}$$
$$s - s\theta = N\theta - s\theta$$
$$\hat{\theta} = \frac{s}{N} = \frac{\sum_{i=1}^{N} x_i}{N}$$

Saç renkleri üzerinden konuşalım



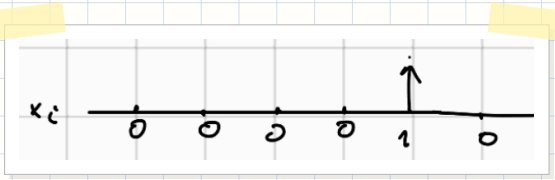
orijinal çizim, s. 2

$$p_k = \left(1 - \sum_{i=1}^{k-1} p_i\right)$$

$$p_i \geq 0, \quad \sum_{i=1}^{k} p_i = 1$$

orj. s. 3

Classlarım var
sınıf vektörü



orijinal çizim, s. 3

← Multinomial Density

$$P(x|\theta) = \prod_{k=1}^{K} p_k^x$$

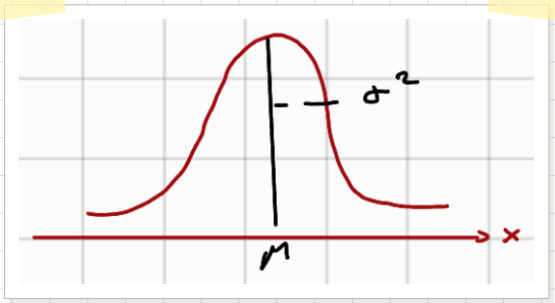
$$L(\theta|\mathcal{X}) = \prod_{i=1}^{N} p(x_i|\theta) = \prod_{i=1}^{N}\prod_{k=1}^{K} p_k^{x_i}$$

$$\ell(\theta|\mathcal{X}) = \log\left(\prod_{i=1}^{N}\prod_{k=1}^{K} p_k^{x_i}\right)$$

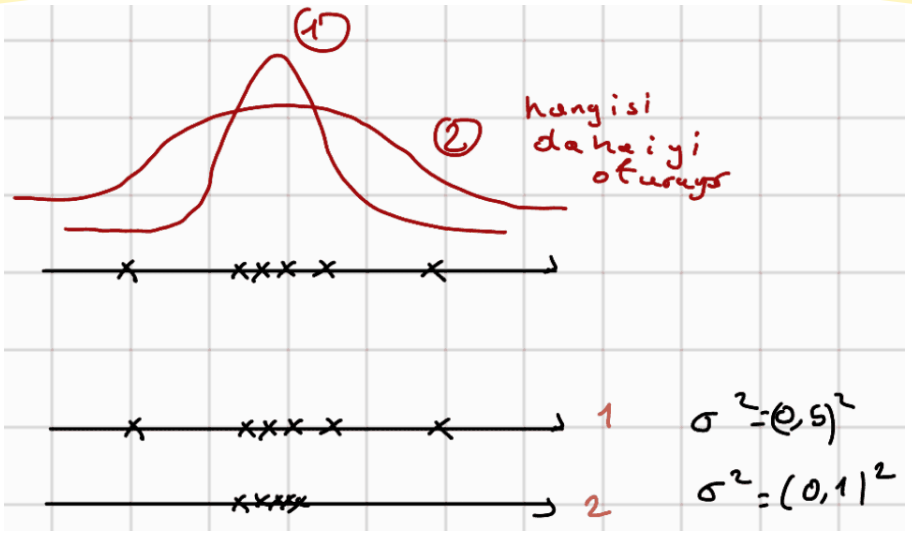
$$\frac{\partial\, \ell(\theta|\mathcal{X})}{\partial p_1} = 0 \Rightarrow \hat{p}_1 = \frac{s_1}{N}$$

Gauss Dağılımı

$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



orijinal çizim, s. 3



orijinal çizim, s. 3

$$L(\theta|\mathcal{X}) = \prod_{i=1}^{N} p(x_i|\theta)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \prod_{i=1}^{N} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\ell(\theta|\mathcal{X}) = -N\log(2\pi\sigma^2) - \sum_{i=1}^{N}\frac{(x-\mu)^2}{2\sigma^2}$$

$$\frac{\partial\,\ell(\theta|\mathcal{X})}{\partial\mu} = \frac{1}{2\sigma^2}\sum_{i=1}^{N} 2(x_i-\mu)$$

$$0 = \frac{\sum_{i=1}^{N} x_i - N\mu}{\sigma^2}$$

orj. s. 4

$$N\cdot\mu = \sum_{i=1}^{N} x_i$$

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\underbrace{\hat{\mu}_{ML}^{T}}_{\text{Maximum likelihood}} = \frac{1}{T}\sum_{t=1}^{T} x_t = \frac{1}{T}\left(x_T + \sum_{t=1}^{T-1} x_t\right) = \frac{x_T}{T} + \frac{1}{T}\sum_{t=1}^{T-1} x_t$$

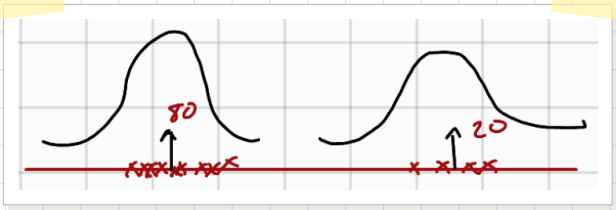
$$= \frac{x_T}{T} + \frac{1}{T}\hat{\mu}_{ML}^{(T-1)}(T-1)$$

$$\hat{\mu}_{ML}^{T} = \frac{1}{T}x_T + \frac{T-1}{T}\hat{\mu}_{ML}^{T-1}$$

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{N}(x_i - \hat{\mu})^2 = 0$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{\mu})^2$$

Gaussian Mixtures

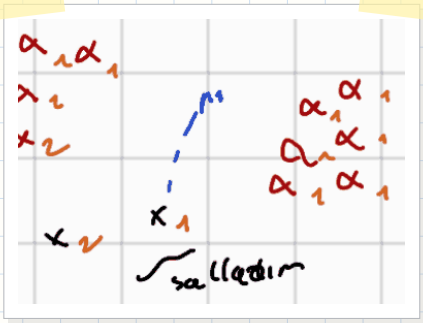


orijinal çizim, s. 4

$$p(x) = \sum_{k=1}^{K} w_k \mathcal{N}(x|\mu_k, \underbrace{\Sigma_k}_{\text{kovaryans matrisi}})$$

k-means:

w_k =?
μ_k =?
Σ_k =?



orijinal çizim, s. 4

EM algoritm

Expectation
maximization algoritm

$$\sum_{k=1}^{K} w_k^0 \mathcal{N}(\mu_k^{(0)}, \Sigma_k^{(0)}) = P(x|\theta)$$

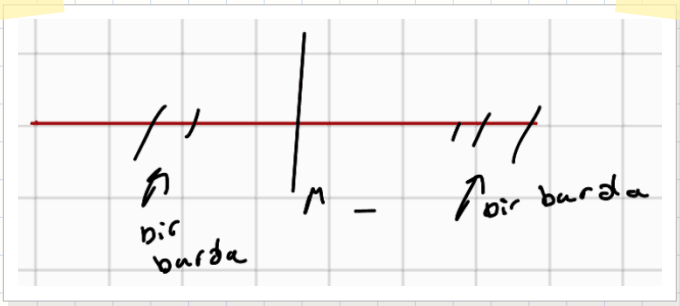
↪ Etiketleme Expectation

↪ Maximization → $\sum_{k=1}^{K} w_k^{(1)} \mathcal{N}(\mu_k^{(1)}, \Sigma_k^{(1)})$

↪ etiket değişmeyene kadar devam ederiz.

orj. s. 5

$$\hat{\mu}^{(N)} = \frac{1}{N} \sum_{i=1}^{N} x_i$$



orijinal çizim, s. 5

buna kestiricinin varyansı

$$\varepsilon(d, \theta) = E\left[(d(x) - \theta)^2\right] \nearrow \text{varyans}$$

$$b_\theta(d) = E\{d(x)\} - \theta$$

$$E\{m\} = E\left\{\frac{\sum_{t=1}^{N} x_t}{N}\right\} = \frac{1}{N} \sum_{t=1}^{N} E\{x^t\}$$

$$= \frac{1}{N} \sum_{t=1}^{N} \mu = \frac{N \cdot \mu}{N} = \mu$$

$$\text{Var}\{m\} = \text{Var}\left\{\frac{\sum_{t=1}^{N} x_t}{N}\right\} = \frac{1}{N^2} \sum_{t=1}^{N} \text{Var}\{x_t\} = \frac{N \cdot \text{Var}\{x_t\}}{N^2}$$

$$= \frac{\text{Var}[x_t]}{N} = \frac{\sigma^2}{N}$$

kestiricinin varyansı

$$s^2 = \frac{\sum_{t=1}^{N} (x_t - m)^2}{N} = \frac{\sum_{t=1}^{N} (x_t)^2 - N m^2}{N}$$

$$E\{s^2\} = \frac{\sum_{t=1}^{N} E\left[(x_t)^2\right] - N E\{m^2\}}{N}$$

$$\text{Var}\{x\} = E\{x^2\} - E\{x\}^2$$

$$E[x^2] = \text{Var}\{x\} + E\{x\}^2 = \sigma^2 + \mu^2$$

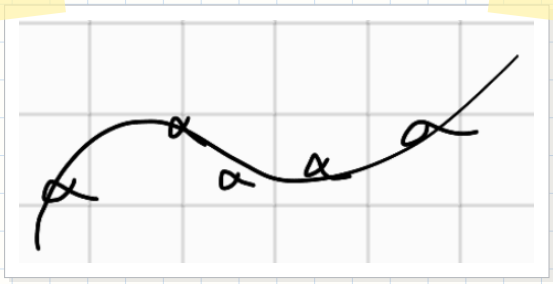
$$E[m^2] = \frac{\sigma^2}{N} + \mu^2$$

orj. s. 6

$$E[s^2] = \left(\sum_{t=1}^{N} \sigma^2 + \mu^2\right) - N \cdot \frac{\sigma^2}{N} - N\mu^2$$

$$= \frac{N\sigma^2 + N\mu^2 - \sigma^2 - N\mu^2}{N} = \frac{(N-1)\sigma^2}{N} \stackrel{?}{=} \sigma^2$$ ↗ yanlı bi sonuç

N ile çarpıp N-1 bölüp, kestiriciyi değiştirsek çıkar

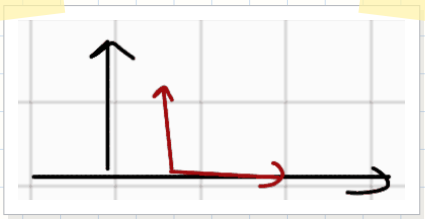


orijinal çizim, s. 6

$$\sum_{i=1}^{N} c_i \underbrace{\left(e^{+j\omega_i k}\right)}_{\text{baz fonksiyonlar}} = f(k)$$

$$f(x) = \sum_{k=1}^{K} w_k \Phi_k(x - x_k)$$

$$\langle f(x), g(x) \rangle = \underbrace{\int_{-\infty}^{\infty} f(x)g(x)\,dx = 0}_{\text{diklik}}$$



orijinal çizim, s. 6

Makine öğrenmesinde dik olmasa da olur base functions

haftaya Naive bayes