

1 Uppgifter Räkneövning 1

- Parameterskattning. Anta att datamängden $\mathcal{X} = \{x_1, \dots, x_N\}$ (obs: x_k är skalär) är slumpat dragna från en normalfördelning med okänt väntevärde m och varians σ^2 .
 - Härled Maximum-likelihood (ML) skattningen av m . (Anta att exemplen är oberoende $\Rightarrow p(\mathcal{X}|m) = p(x_1|m)p(x_2|m)\dots p(x_N|m)$.)
 - Anta att *a priori*-fördelningen för m är Gaussisk med väntevärde m_0 och varians σ_0^2 .¹ Härled maximum *a posteriori* (MAP) skattningen av m .
- Upprepa uppgift 1 (a) och (b) under generaliseringen att $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, dvs exemplen är vektorer istället för skalärer. Anta $\mathbf{x} \sim N(\mathbf{m}, \mathbf{C})$ och $\mathbf{m} \sim N(\mathbf{m}_0, \mathbf{C}_0)$ (i MAP skattningen).
- Algorithm för uppdatering av parametrarna i Gaussiska mixturer. Anta att vi har en datamängd, $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, där exemplen dragits från Gaussisk mixtur med M st komponenter:

$$p(\mathbf{x}_n|\theta) = \sum_m^M p(\mathbf{x}_n|\mathbf{m}_m, \mathbf{C}_m)P_m, \quad (1)$$

där mängden $\theta = \{\mathbf{m}_1, \dots, \mathbf{m}_M, \mathbf{C}_1, \dots, \mathbf{C}_M, P_1, \dots, P_M\}$ representerar de samlade okända parameterarna. PDFen för den m te komponenten, $p(\mathbf{x}|\mathbf{m}_m, \mathbf{C}_m)$, ges av

$$p(\mathbf{x}|\mathbf{m}_m, \mathbf{C}_m) = \frac{1}{(2\pi)^{d/2}|\mathbf{C}_m|^{d/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m}_m)^T \mathbf{C}_m^{-1}(\mathbf{x} - \mathbf{m}_m)\right) \quad (2)$$

där $d = \dim(\mathbf{x})$.

- Visa att ML-skattning av väntevärdesvektorerna, $\mathbf{m}_1, \dots, \mathbf{m}_M$, leder till att vi behöver lösa ekvationssystemet

$$\hat{\mathbf{m}}_k = \frac{\sum_n P(k|\mathbf{x}_n)\mathbf{x}_n}{\sum_m P(k|\mathbf{x}_m)} \quad (3)$$

för $k = 1, \dots, M$ där $P(k|\mathbf{x}_n)$ i sin tur ges av:

$$P(k|\mathbf{x}_n) = \frac{p(\mathbf{x}_n|\hat{\mathbf{m}}_k, \mathbf{C}_k)P_k}{\sum_{m=1}^M p(\mathbf{x}_n|\hat{\mathbf{m}}_m, \mathbf{C}_m)P_m}. \quad (4)$$

- På liknande sätt, visa att följande ekvationer måste vara uppfyllda för $\hat{\mathbf{C}}_k$:

$$\hat{\mathbf{C}}_k = \frac{\sum_n P(k|\mathbf{x}_n)(\mathbf{x}_n - \hat{\mathbf{m}}_k)(\mathbf{x}_n - \hat{\mathbf{m}}_k)^T}{\sum_m P(k|\mathbf{x}_m)} \quad (5)$$

- EM-algoritmen för identifiering av Gaussmixturer och KMC-algoritmen har vissa likheter. Genom att fixera kovariansmatriserna till $\mathbf{C}_1 = \dots = \mathbf{C}_M = \mathbf{I}$, fixera $P_1 = \dots = P_M = 1/M$ samt göra vissa lämpliga approximationer av $P(k|\mathbf{x}_n)$ i ekv. (4) så blir EM-algoritmen identisk med KMC. Vilka approximationer handlar det om?

¹Denna *a priori*-fördelning kan tex baseras på en tidigare skattning med en annan datamängd.

5. Visa att både Histogramskattning och kärnskattning (Parzen windows) ger PDF-skattningar som alltid integrerar till 1.
6. Visa att PDF-skattning m.h.a Gauss-mixtur alltid integrerar till 1.
7. Visa med ett enkelt exempel att PDF-skattningar erhållna med KNN *inte* integrerar till 1.
8. Sannolikhet sett som ett väntevärde. En sannolikhet kan ses som ett väntevärde av en s.k. *indikatorfunktion*. För att illustrera detta betrakta ett två klass-problem med klasserna C_1 och C_2 . Låt $t(\mathbf{x})$ vara en indikatorvariabel som antar värdet "1" om mönstret dragits från C_1 och "0" om det dragits från C_2 . Vi kan se t som en stokastisk variabel som beror av \mathbf{x} .
 - (a) Utgå från definitionen på väntevärde för att visa att $E[t] = P(C_1)$.
 - (b) På liknande sätt, visa att $E[t|\mathbf{x}] = P(C_1|\mathbf{x})$
9. Förberedelse för modellparameterskattning. Betrakta den enkla modellen

$$y = wx + e_y, \tag{6}$$

där e_y antas vara normalfördelad, $e_y \sim N(0, \sigma_y^2)$. Anta att vi har en *a priori* normalfördelning på vikten w , $w \sim N(0, \sigma_w^2)$. Vår datamängd består av $\mathbf{Y} = (y_1, y_2, \dots, y_N)$ samt $\mathbf{X} = (x_1, x_2, \dots, x_N)$ (\mathbf{X} och \mathbf{Y} är radvektorer). Vilka är ML respektive MAP skattningarna av w ? Visa dessutom att ML skattningen erhålls som ett specialfall av MAP skattningen då $\sigma_w^2 \rightarrow \infty$.