

Räkneövning 4

1. Algoritm 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 $d = \dim(\mathbf{x})$.

- (a) Visa att ML-skattningar av väntevärdesvektorer, $\mathbf{m}_1, \dots, \mathbf{m}_M$, måste uppfylla 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:

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

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

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

2. KMC kan ses som en "fattigmansversion" av EM-algoritmen som ges nedan:

$$\mathbf{m}_j^{new} = \frac{\sum_n P^{old}(j|\mathbf{x}_n)\mathbf{x}_n}{\sum_n P^{old}(j|\mathbf{x}_n)} \quad (6)$$

$$(\sigma_j^{new})^2 = \frac{1}{d} \frac{\sum_n P^{old}(j|\mathbf{x}_n) \|\mathbf{x}_n - \mathbf{m}_j^{new}\|^2}{\sum_n P^{old}(j|\mathbf{x}_n)} \quad (7)$$

$$\mathbf{m}_j^{new} = \frac{1}{N} \sum_n P^{old}(j|\mathbf{x}_n) \mathbf{x}_n \quad (8)$$

Skriv explicit ut de förenklingar och approximationer i EM-algoritmen som leder till KMC.

3. Hierarkisk klustring, allmänt.

- (a) Skriv upp huvudalgoritmen för hierarkisk klustring.

- (b) Ange tre olika sätt att mäta avstånd mellan kluster. Namnge avståndsmåtten.
 - (c) Vad är ett dendrogram? Vilken information kan man utläsa ur ett dendrogram.
4. Hierarkisk klustring, tekniska detaljer.
- (a) Anta att du ska implementera någon av varianterna för hierarkisk klustring och att dina mönstervektorer har hög dimension. Den "råa" algoritmen för hierarkisk klustring (den som vi gick igenom på föreläsningen) inbegriper ett stort antal jämförelser mellan olika vektorer (vi kan tvingas räkna ut ett stort antal avstånd $d(\mathbf{x}_n, \mathbf{x}_k)$.) Ge ett förslag på hur beräkningarna kan organiseras så att onödiga beräkningar kan undvikas.
5. UFAT
- (a) Försök att med hjälp av en figur förklara varför två lager är tillräckligt för att en flerlagerperceptron (med tillräckligt många neuroner i första lagret och linjärt sista lager) är en universell funktionsapproximator. Gör detta för fallet då nätet skall approximera en funktion av *en* variabel.
 - (b) Försök att med hjälp av en figur förklara varför en RBF (med tillräckligt många basfunktioner) också är en universell funktionsapproximator. Återigen, betrakta approximation av en funktion av *en* variabel.
6. Säg att vi har en urna som innehåller bollar som kan vara antingen röda eller vita. Förutom färgskillnaden så är bollarna helt identiska. Vi vet att urnan innehåller exakt 50 röda, och att antalet vita kan vara vad som helst mellan 1 och 100 (gränserna inkluderade). Vi står i färd att dra en boll från urnan. Vad är sannolikheten att den är röd?
7. Skattning av sannolikheter: "Bayesiansk variant". Betrakta en slant som vi singlar ett antal gånger och ser hur ofta vi får krona respektive klave. Med hjälp av detta vill vi uppskatta sannolikheten för krona och klave vid nästa kast. Anta att vi just den här gången har fått slanten av en bekant som jobbar på Ångströmlabbet. Hans specialitet är legeringar och han är "en riktig filur". Detta gör att vi absolut inte kan anta att myntet (som troligtvis är en förfalskning) är balanserat och (vilket skulle ge samma sannolikhet för krona och klave). Låt $\theta = P('krona')$ beteckna (den okända) sannolikheten för krona vilket ger oss $1 - \theta = P('klave')$. Vi sammanfattar vår väldigt osäkra bakgrundsinformation genom a priori pdf:en $p(\theta) = 1$, dvs vi tror från början lika mycket på alla tänkbara värden på θ .
- (a) Nu observerar vi en sekvens av kast: $S = \{\text{krona, krona, krona}\}$. Hur har vår uppfattning om θ ändrats genom denna nya information? dvs hur ser $p(\theta|S)$ ut? Plotta den.
 - (b) Stämmer intuitionen här? Vad säger pdf:en tex om hypotesen " $\theta = 0$ "?
 - (c) Räkna ut sannolikheten att θ ligger i intervallet $[0.45, 0.55]$.
 - (d) Vilken är ML-skattningen?

- (e) Anta att vi nu tänker kasta myntet ännu en gång. Vilken är sannolikheten för krona vid detta kast enligt det Bayesianska synsättet?
8. Anta att du vill beräkna sannolikheten att en HMM har befunnit sig i ett visst tillstånd, q_i , vid tiden t , dvs vi vill beräkna $P(s^t = q_i | X, \theta)$, där X är sekvensen $X = \{X^1, X^2, \dots, X^T\}$ och θ står för våra modellparametrar. Om vi utnyttjar Bayes sats får vi

$$P(s^t = q_i | X, \theta) = \frac{P(X, s^t = q_i | \theta)}{P(X | \theta)} \quad (9)$$

- (a) Visa att uttrycket $P(X, s^t = q_i | \theta)$ går att utveckla som

$$P(X, s^t = q_i | \theta) = \beta_i(t) \alpha_i(t), \quad (10)$$

där $\alpha_i(t)$ är den variabel vi arbetar med i framåtalgoritmen och $\beta_i(t) = P(X^{t+1}, \dots, X^T | s^t = q_i, \theta)$.

- (b) Hur kan vi beräkna nämnaren?