Räkneövning 1

1. Anta att vi har två stycken tärningar, T_1 och T_2, vilka kan beskrivas av sannolikheterna i tabellen nedan.

T_1	T_2
$P(1) = 1/6$	$P(1) = 0.1$
$P(2) = 1/6$	$P(2) = 0.1$
$P(3) = 1/6$	$P(3) = 0.2$
$P(4) = 1/6$	$P(4) = 0.2$
$P(5) = 1/6$	$P(5) = 0.2$
$P(6) = 1/6$	$P(6) = 0.2$

Table 1: Utfallssannolikheter för tärning T_1 och T_2

   (a) Anta att någon av tärningarna slås (du vet inte vilken) och utfallet blir 5. Beräkna sannolikheten för att tärningskastet gjordes med T_1 respektive T_2.

   (b) Anta att någon av tärningarna slås 15 ggr med utfallen:
   $\{3, 4, 6, 6, 6, 5, 3, 4, 5, 2, 3, 6, 5, 4, 5\}$. Vilka är nu sannolikheterna för att det är T_1 respektive T_2?

   (c) Vilken är den allmänna formeln för *a posteriori* sannolikheten då vi observerar en sekvens bestående av totalt n_1 ettor, n_2 tvåor osv.?

2. Skattning av sannolikheter. Tänk er en oregelbunden tärning med L st. sidor som vi kastar ett stort antal gånger. Vi är intresserade av att uppskatta sannolikheterna för att tärningen hamnar med olika sidorna upp. Vi observerar sekvensen av utfall $O = \{X_1, ..., X_N\}$ och räknar ihop antalet ettor till n_1, antal tvåor till n_2, osv. Vi betecknar de respektive sannolikheterna för $P_1, ..., P_L$ och dessa är de okända parametrarna som skall skattas.

   (a) Vilken är likelihoodfunktionen i detta fall?

   (b) Vilket bivillkor måste vara uppfyllt för värdena på $P_1, ..., P_L$?

   (c) Ställ upp Lagrangefunktionen, $L(P_1, ..., P_L, \lambda)$, för optimeringsproblemet och visa att ML-skattningen av sannolikheterna ges av $P_i = n_i/N$.

3. ML-skattningarna av tärningssannolikheterna i föregående uppgift är troligen noggranna om antalet kast är mycket stort. De skattade sannolikheterna kan vi nu använda i en modell för tärningen och utnyttja den vid jämförelse med andra modeller (jmfr uppgift 2). Anta nu att vi istället hade tillgång till endast ett fåtal observationer. Kan du se några potentiella problem med att använda skattningen ovan. Ledning: Välj att betrakta något extremfall.

4. I vissa mönsterigenkänningsproblem har man möjligheten att lämna vissa mönster oklassificerade. Låt klassificeringskostnaden för korrekt val av klass vara noll, λ_r för

att lämna oklassificerade (rejected) och λ_s när klassningen är fel, dvs

$$\lambda(a_i|\omega_j) = \begin{cases} 0 & i = j \\ \lambda_r & i = M+1 \\ \lambda_s & \text{annars} \end{cases} \quad (1)$$

där a_i representerar beslutet "klass i". Vi har härigenom introducerat en handling (action) a_{M+1} (förutom de övriga sorteringarna) och λ_r är kostnaden att välja denna handling.

(a) Visa att den optimala beslutsregeln som minimerar den förväntade risken kan uttryckas som:
- Välj klass ω_k om $P(\omega_k|x) \geq P(\omega_i|x)$ för $i \neq k$ och $P(\omega_k|x) \geq (1 - \frac{\lambda_r}{\lambda_s})$
- Annars, lämna oklassificerad (välj "klass" ω_{M+1}).

(b) Vad händer vid specialfallen $\lambda_r = 0$ and $\lambda_r > \lambda_s$.

5. Betrakta ett klassificeringsproblem med två klasser, ω_1 och ω_2. De har *a priori* sannolikheterna $P(\omega_1) = P_1$ respektive $P(\omega_2) = P_2$ och är normalfördelade $N(\mathbf{m}_1, \mathbf{C}_1)$ respektive $N(\mathbf{m}_2, \mathbf{C}_2)$. Vårt mål är att klassificera en mönstervektor $\mathbf{x}$ och vi väljer att minimera sannolikheten för fel beslut.

(a) Ange en lämplig "loss matrix" för detta problem.

(b) Ställ upp lämpliga diskriminantfunktioner för klasserna, $g_1(\mathbf{x})$ och $g_2(\mathbf{x})$. (De ska vara explicit uttryckta i de parametrar $P_1, P_2, \mathbf{m}_1$ osv som anges ovan.

(c) Beslutsytan ges av lösningen till ekvationen $g_1(\mathbf{x}) = g_2(\mathbf{x})$. Ställ upp och förenkla så långt det går för följande specialfall:

i. Kovariansmatriserna är lika, dvs $\mathbf{C}_1 = \mathbf{C}_2 = \mathbf{C}$. Rita ett illustrerande exempel.

ii. Kovariansmatriserna är lika och "cirkulära", dvs $\mathbf{C}_1 = \mathbf{C}_2 = \sigma^2\mathbf{I}$ där $\mathbf{I}$ är enhetsmatrisen. Rita exempel.

(d) Om $\mathbf{x} = (1,1)^T$, $\mathbf{m}_1 = (0,0)^T$,

$$\mathbf{C}_1 = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \quad (2)$$

vad är det s.k. *Mahalanobisavståndet* mellan $\mathbf{x}$ och $\mathbf{m}_1$?

6. Parameterskattning. Anta att datamängden $\mathcal{X} = \{x_1, ..., x_N\}$ (obs: x_k är skalär) är slumptal dragna från en normalfördelning med okänt väntevärde m och varians σ^2.

(a) Härled Maximum-likelihood (ML) skattningen av m. (Anta att exemplen är oberoende $\Rightarrow p(\mathcal{X}|m) = p(x_1|m)p(x_2|m)...p(x_N|m)$.)

(b) 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.

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

7. Upprepa uppgift 6 (a) och (b) under generaliseringen att $\mathcal{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$, 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).

8. Algoritm för uppdatering av parametrarna i Gaussiska mixturer. Anta att vi har en datamängd, $\mathcal{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$, där exemplen dragits från Gaussisk mixtur med M st komponenter:

$$p(\mathbf{x}_n|\theta) = \sum_m^M p(\mathbf{x}_n|\mathbf{m}_m, \mathbf{C}_m)P_m, \tag{3}$$

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

$$p(\mathbf{x}|\mathbf{m}_m, \mathbf{C}_m) = \frac{1}{(2\pi)^{d/2}|\mathbf{C}_m|^{d/2}} exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m}_m)^T\mathbf{C}_m^{-1}(\mathbf{x} - \mathbf{m}_m)\right) \tag{4}$$

där $d = dim(\mathbf{x})$.

(a) Visa att ML-skattningar av väntevärdesvektorerna, $\mathbf{m}_1, ..., \mathbf{m}_M$, måste uppfylla ekvationssystemet

$$\hat{\mathbf{m}}_k = \frac{\sum_n P(k|\mathbf{x}_n)\mathbf{x}_n}{\sum_m P(k|\mathbf{x}_m)} \tag{5}$$

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

$$P(k|\mathbf{x}_n) = \frac{p(\mathbf{x}_n|\hat{\mathbf{m}}_k, \mathbf{C}_k)P_k}{\sum_{m=1}^M p(\mathbf{x}_n|\hat{\mathbf{m}}_m, \mathbf{C}_m)P_m}. \tag{6}$$

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

$$\hat{\mathbf{C}}_k = \frac{\sum_n P(k|\mathbf{x}_n)(\mathbf{x}_n - \hat{\mathbf{m}}_k)(\mathbf{x}_n - \hat{\mathbf{m}}_k)^T}{\sum_m P(k|\mathbf{x}_m)} \tag{7}$$