

Räkneövning 3

1. Två klasser har för samma egenskap värdena $\{1, 2, 3, 4, 5, 6, 7\}$ (klass ω_1) och $\{4, 5, 6, 7, 8, 9, 10\}$ (klass ω_2). Utnyttja t-testet för att avgöra om denna egenskap är värd att använda vid klassificering. Välj signifikansnivån $\rho = 2.5\%$.
2. Antag att vi har en diskret fördelning där data genereras med samma sannolikhet från fyra olika punkter. Låt oss studera två fall
I: punkterna är $(1,1)$, $(2,2)$, $(-1,-1)$ och $(-2,-2)$.
II: punkterna är $(1,-1)$, $(2,2)$, $(-1,1)$ och $(-2,-2)$.
 - (a) Rita två figurer där punkterna för respektive (punkt-)fördelning är utritade.
 - (b) Bestäm fördelningarnas väntevärden.
 - (c) Bestäm kovariansmatriserna.
 - (d) Bestäm egenvektorer till respektive kovariansmatris och konstruera ny ON-bas.
 - (e) Rita in de nya basvektorerna i figuren från deluppgift (a)
 - (f) Uttryck de fyra diskreta punkterna i den nya basen.
 - (g) Plocka bort den komponent som svarar mot minsta egenvärdet. Hur stort blir approximationsfelet i medeltal då man representerar data endast med hjälp av en egenvektor i fallen I och II?
3. PCA baseras på matrisen $\mathbf{A} = (\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_M)$ där $\mathbf{a}_i$ är den i :te egenvektorn till datakovariansmatrisen $\mathbf{C}$. Ordningen mellan vektorerna ges av $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$ där λ är egenvärdena till $\mathbf{C}$. Med hjälp av $\mathbf{A}$ kan vi nu bilda en ny vektor $\mathbf{y} = \mathbf{A}^T(\mathbf{x} - \mathbf{m})$ samt en approximation av $\mathbf{x}$ som vi skriver $\hat{\mathbf{x}} = \mathbf{A}\mathbf{A}^T(\mathbf{x} - \mathbf{m}) + \mathbf{m}$.
 - (a) Visa att komponenterna i vektorn $\mathbf{y}$ är okorrelerade.
 - (b) Visa att variansen för varje komponent ges av motsvarande egenvärde, dvs att $E[(y_i - m_y^i)^2] = \lambda_i$, där $m_y^i = E[y_i]$.
 - (c) Låt $\mathbf{e}$ beteckna rekonstruktionsfelet, dvs
$$\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}. \tag{1}$$Visa att förväntade kvadratfelet $J = E[||\mathbf{e}||^2] = \sum_{k=M+1}^d \lambda_k$, där $d = \dim(\mathbf{x})$.
Ledtråd: Ni kan behöva utnyttja sambandet mellan $\text{tr}(\mathbf{C})$ och egenvärdena till $\mathbf{C}$.
 - (d) Visa att resultatet i uppgift (c) är rimligt genom att verifiera att resultatet är korrekt för ytterligheterna $M = 0$ och $M = d$.
4. Fisher index. Anta att vi vill finna den vektor, $\hat{\mathbf{w}}$, som maximerar Fisher index då data (2-d) kommer från två klasser, ω_1 och ω_2 , med väntevärden $\mathbf{m}_1$ respektive $\mathbf{m}_2$ samt med kovariansmatriserna $\mathbf{C}_1$ respektive $\mathbf{C}_2$. Om vår träningsmängd innehåller många exempel dragna från fördelningarna för ω_1 och ω_2 så får du anta att $\mathbf{S}_w$ kan skrivas med stor noggrannhet som $\mathbf{S}_w = n_1\mathbf{C}_1 + n_2\mathbf{C}_2$, där n_1 och n_2 är antalet exempel från varje klass. Låt $\mathbf{m}_1 = (1,1)^T$ och $\mathbf{m}_2 = (0,0)^T$. Räkna ut $\hat{\mathbf{w}}$ i det fall att

- (a) både ω_1 och ω_2 har kovariansmatrisen $\sigma^2 \mathbf{I}$. Rita figur!
- (b) både ω_1 och ω_2 har samma diagonala kovariansmatris $\mathbf{D}$. Diagonalelementen är d_1^2 och d_2^2 . Rita figur! (Hur viktas de olika komponenterna i skillnadsvektorn $(\mathbf{m}_1 - \mathbf{m}_2)$ för olika val av d_1 och d_2).
5. Projektionerna på vektorerna vi räknat fram via linjär diskriminantanalys ger ingen automatiskt klassning. För detta måste vi bestämma även ett tröskelvärde, θ , och utnyttja beslutsregeln: välj ω_1 om $\mathbf{w}^T \mathbf{x} > \theta$, annars välj ω_2 . Optimal klassning (minimal felsannolikhet) för specialfallet normalfördelade klasser med samma kovariansmatriser har ett tydligt släktskap med Fishers linjära diskriminant och för detta specialfall är har vi ett explicit uttryckt θ . Red ut släktskapet och ge uttrycket för θ .
6. Följdfråga: gäller det omvända till slutsatsen i förra uppgiften? Det vill säga, om Fishers diskriminant för två klasser råkar sammanfalla med den optimala beslutsgränsen, kan vi då dra slutsatsen att data är normalfördelade?
7. Säg att vi för våra exempel $\mathbf{x}_1, \dots, \mathbf{x}_N$ (från två klasser, ω_1 och ω_2) har räknat ut projektioner $y_1, \dots, y_N$ m.h.a. Fishers projektionsvektor, $\mathbf{w} = \mathbf{S}_w^{-1}(\mathbf{m}_1 - \mathbf{m}_2)$. Anta att någon annan person har fått nästan samma data som du, enda skillnaden är att dessa data är transformerade med en inverterbar matris $\mathbf{A}$, dvs exemplen ges av $\tilde{\mathbf{x}}_n = \mathbf{A}\mathbf{x}_n$.
- (a) Visa att ni kommer att få samma y -värden.
- (b) Slutsatser? Ledning: Behövs normering av koordinaterna i $\mathbf{x}$ -vektorerna före uträknad av projektionerna?
8. Linjär regression och regularisering. Vi har träningsexemplen $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ med tillhörande y -värden $\{y_1, \dots, y_N\}$ och vi finna en vektor $\hat{\mathbf{w}}$ så att vi kan approximera y_n med $y_n = \hat{\mathbf{w}}^T \mathbf{x}_n$.
- (a) Härled ett uttryck för den vektor $\hat{\mathbf{w}}$ som minimerar kvadratsummekriteriet $J = \sum_{n=1}^N (y_n - \hat{\mathbf{w}}^T \mathbf{x}_n)^2$.
- (b) Härled ett uttryck för den vektor $\hat{\mathbf{w}}$ som minimerar regulariserade kvadratsummekriteriet $J = \sum_{n=1}^N (y_n - \hat{\mathbf{w}}^T \mathbf{x}_n)^2 + \lambda \|\mathbf{w}\|^2$.
9. Vi har träningsexemplen $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ med tillhörande y -värden $\{y_1, \dots, y_N\}$ som vi antar har uppkommit från den stokastiska modellen

$$y_n = \mathbf{w}^T \mathbf{x}_n + e_n \quad (2)$$

där e_n är normalfördelad $e_n \sim N(0, \sigma^2)$.

- (a) Härled maximum likelihood (ML) skattningen av parametervektorn $\mathbf{w}$.
- (b) Anta att vi vet *a priori* att $\mathbf{w}$ troligtvis består av relativt små värden. Vi modellerar detta genom att säga att $\mathbf{w}$ *a priori* är normalfördelad med väntevärde noll och med en kovariansmatris $\mathbf{C}_w = \sigma_w^2 \mathbf{I}$ där $\mathbf{I}$ är enhetsmatrisen och variansen σ_w^2 anger ett mått på vad vi menar med "små värden". Härled maximum *a posteriori* (MAP) skattningen av $\mathbf{w}$.
- (c) Jämför svaren i (a) och (b) med svaren på fråga ??.