

Uppsala universitet
Signaler and System
MG 018 - 471 3229

Tentamen i datoriserad mönsterigenkänning

för civilingenjörsprogrammen i teknisk fysik och informationsteknologi samt för matematiskt naturvetenskapligt program.

Plats och Tid: Post Scriptum 2, onsdag 9 januari, 2002, 9.00-14.00

Hjälpmedel: Formelsamling på kursens hemsida.

Obs! Svar utan motivering ger noll poäng!

Problem #1 (6p=3p+2p+1p)

Antag att sannolikhetstäthetsfunktionerna för två lika troliga klasser ω_1 och ω_2 ges av

$$p(\mathbf{x}|\omega_1) = \begin{cases} 1/9 & \text{om } 1 \leq x_1 \leq 4 \text{ och } 1 \leq x_2 \leq 4 \\ 0 & \text{annars} \end{cases}$$

och

$$p(\mathbf{x}|\omega_2) = \begin{cases} 1 & \text{om } 2 \leq x_1 \leq 3 \text{ och } 2 \leq x_2 \leq 3 \\ 0 & \text{annars} \end{cases}$$

- Rita beslutsgränser för den optimala klassificerare som minimerar antalet klassificeringsfel (i medeltal).
- Bestäm den optimala klassificerare som minimerar antalet klassificeringsfel (i medeltal).
- Beräkna klassificeringsfelet för den optimala klassificeraren som minimerar antalet klassificeringsfel (i medeltal).

Problem #2 (6p=3p+2p+1p)

- Civ har 40 designexempel och vill bygga en linjär regressionsmodell. Hon bygger med hjälp av OLS två olika modeller, en med hjälp av de 20 första exemplen, en annan med hjälp av de 20 sista. Modellerna blir dock helt olika! Ge en trolig förklaring till detta fenomen och ge förslag på två metoder som kanske kan lösa detta problem. Förklara också med några få meningar vad det är som gör att dessa två metoder kanske är bättre än OLS i detta fall.
- Förklara hur man stegvis designar en linjär klassificerare med hjälp av minsta kvadratmetoden. Antag att de två klasser som man försöker designa en linjär

klassificerare för är linjärt separabla. Varför är det troligen bättre att använda perceptronregeln (istället för minsta kvadratmetoden) i detta fall?

c) Motivera med hjälp av en figur ekvationen för ett hyperplan med normalvektor $\mathbf{n}$ som passerar genom punkten $\mathbf{x}_o$.

Problem #3 (6p=3p+2p+1p)

Antag att N fyrdimensionella vektorer $\mathbf{x}_i, i = 1, 2, 3, \dots$ ska överföras via en smalbandig teleledning. På grund av den begränsad kapaciteten måste varje vektor reduceras till två (reella) tal innan överföring (vi bortser här från problemet med binär representation av reella tal). Med hjälp av ett antal designexempel estimeras kovariansmatrisen för de fyrdimensionella vektorerna och resultatet är egenvektorerna $\mathbf{w}_1 = [0.78, -0.61, 0, 0]^T$, $\mathbf{w}_2 = [0.61, 0.78, 0, 0]^T$, $\mathbf{w}_3 = [0, 0, 1, 0]^T$ och $\mathbf{w}_4 = [0, 0, 0, 1]^T$ med egenvärden $\lambda_1 = 0.44$, $\lambda_2 = 4.56$, $\lambda_3 = 4.0$, $\lambda_4 = 1.0$. OBS! Egenvektorerna är inte ordnade efter storleken på egenvärdena!

a) Antag att den vektorn $\mathbf{x} = [1, 0, 1, 0]^T$ står på tur att överföras. Bestäm de två tal som får representera vektorn $\mathbf{x}$ vid överföringen så att rekonstruktionen på mottagarsidan (där man har tillgång till alla egenvektorer och egenvärden) blir så bra som möjligt (i minsta kvadratmening). Bestäm också hur den rekonstruerade vektorn $\hat{\mathbf{x}}$ kommer att se ut på mottagarsidan. Bestäm till sist också kvadraten på rekonstruktionsfelet: $\|\mathbf{x} - \hat{\mathbf{x}}\|^2$ för denna vektor.

b) Förklara med hjälp av storleken på egenvärdena varför det är väntat att rekonstruktionen i deluppgift a) blev så dålig. Förklara också varför resultatet inte skulle bli speciellt bra även om man tillät att tre istället för två tal får överföras för varje bild.

c) Varför kan man (under antagande om normalfördelning) säga att det är ganska oväntat att man ska observera vektorn $\mathbf{x} = [1, 0, 1, 0]^T$ dvs att denna vektor nästan kan betraktas som en outlier?

Problem #4 (6p=3p+2p+1p)

a) Antag att den betingade sannolikhetsfördelningen $P(D|k)$ för en datamängd D givet en okänd heltalsparameter k kan skrivas

$$P(D|k) = \begin{cases} 1/3 & \text{om } k = 1 \\ 1/2 & \text{om } k = 2 \\ 1/6 & \text{om } k = 3 \\ 0 & \text{om } k = 4 \end{cases}$$

Antag vidare att sannolikhetsfördelningen för k apriori definieras av

$$P(k) = \begin{cases} 0.33 & \text{om } k = 1 \\ 0.33 & \text{om } k = 2 \\ 0.01 & \text{om } k = 3 \\ 0.33 & \text{om } k = 4 \end{cases}$$

Bestäm följande två estimat av parametern k : maximum likelihood (ML) och maximum a posteriori (MAP). Beräkna också $E\{k|D\}$.

b) Härled stationära villkoren dvs normalekvationerna för parametrarna i en linjär regressionsmodell då man har tillgång till ett ändligt respektive ett oändligt antal designexempel och använder medelkvadratskriteriet. Med oändligt antal menas här tillgång till sannolikhetstäthetsfunktionerna.

c) Visa att $g(x) = E\{y|x\}$ minimierar kriteriet $J(g) = \int_x \int_y (y-g(x))^2 p(x,y) dy dx = \int_x \int_y (y-g(x))^2 p(y|x)p(x) dy dx$.

Problem #5 (6p=3p+2p+1p)

a) Visa att funktionen $J(m) = \sum_{n=1}^N (x_n - m)^2$ minimieras av aritmetiska medelvärdet $m^* = \frac{1}{N} \sum_{n=1}^N x_n$.

b) Givet N exempel från en stokastisk variabel, förklara varför medianen är en mer robust approximation av väntevärdet än det aritmetiska medelvärdet. Förklara också vad detta har för betydelse vid klassificering med hjälp av optimala Bayesianska klassificerare i specialfallet normalfördelade variabler.

c) Visa att funktionen $J(m) = \sum_{n=1}^N |x_n - m|$ där $|x|$ betyder absolutbeloppet av x och N är ett udda tal minimieras av medianen för alla x_n , $m^* = \text{median}(x_1, x_2, \dots, x_N)$. Ledning: Dela först upp summan i två delsummer med hjälp av definitionen av $|x|$.

Problem #6 (6p=3p+3p)

Civ ska bygga ett system för igenkänning av ord som registreras via en mikrofon från olika användare. Civ kallar in en person som läser in tre olika ord. Varje ord läses flera gånger och lagras varje gång i form av en vektor bestående av 20000 sampel (värden) från mikrofonsignalen motsvarande ett tidsintervall med längden 1 sekund.

a) Civ försöker bygga en klassificerare baserat direkt på kNN-metoden och Euklidiskt avstånd. Förklara med hjälp av ord och figurer och i termer av kNN-metoden varför Civ kommer att få stora problem om personen som läser in i) inte börjar tala vid exakt samma tidpunkt varje gång och/eller ii) varierar ljudstyrkan kraftigt mellan olika inläsningar och/eller iii) varierar längden på uttalet mellan olika inläsningar.

b) Varje sampel lagras med 8 bitars upplösning dvs 256 olika nivåer. Civ konverterar nu varje inläst mikrofonsignal till ett histogram över dessa 256 nivåer och använder detta histogram som mönstervektor istället. Motivera minst två tydliga fördelar och en eventuell nackdel med denna metod i jämförelse med att använda de ursprungliga 20000-dimensionella mönstren.

Problem #7 (6p=3p+3p)

a) Perceptronregeln kan tolkas som en iterativ metod som försöker minimera kriteriet $J(\mathbf{w}) = \sum_{\mathbf{x}_i \in \mathcal{M}} (-\mathbf{w}^T \mathbf{x}_i)$ där $\mathcal{M} = \{\mathbf{x}_i | \mathbf{w}^T \mathbf{x}_i < 0\}$. Förklara, vad är $\mathbf{w}$, och $\mathcal{M}$ i detta sammanhang och varför kallas perceptronregeln för en felkorrigering metod?

b) Minsta kvadratmetoden är ett alternativ till perceptronregeln som normalt bygger på lösningen till ett ekvationssystem men som också kan lösas iterativt på samma sätt som perceptronregeln. Härled perceptronregeln och sedan också motsvarande iterativa algoritmen som minimerar minsta kvadratkriteriet.

Problem #8 (6p=3p+3p)

a) K-närmaste grannemetoden (kNN) är en intuitiv och ofta kraftfull metod för klassificering. Vilka är de två huvudsakliga "parametrarna" som man kan variera i denna metod för att få olika resultat? Förklara också varför det inte alltid är bäst att klassa baserat enbart på närmaste grannen ($K=1$).

b) Antag att det finns C klasser och en prototypvektor $\mathbf{p}_c$ för varje klass $c = 1, 2, \dots, C$. Antag vidare att närmaste granne metoden ska användas, $K = 1$, och att Euklidiskt avstånd mäter likhet. Visa att denna klassificerare kan implementeras i form av C linjära diskriminantfunktioner.

Lycka till!