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: Polacksbacken, onsdag 23 april 2003, 9.00-14.00

Hjälpmedel: Räknare, Mathematical Handbook, Matehmatics Hanndbok (BETA), formelsamling på kursens hemsida.

Betygsgränser: 3: 18p 4:26p 5:32p

Obs! Svar utan motivering ger noll poäng!

Problem #1 (7p=3p+2p+2p)
a) Förklara i ett antal punkter hur man med hjälp av ett antal exempel designar och verifierar ett fungerande mönsterigenkänningssystem "från ax till limpa" dvs hela vägen från datansamling till slutvalidering).
b) Vad är fördelen respektive nackdelen med att använda en testmängd i jämförelse med att utföra korsvalidering?
c) Antag att det okända verkliga (optimala,Bayesianska) felet är e_T och att bland N oberoende test exempel blir k stycken felklassade. Detta inträffar med sannolikheten

$$P(k|e_T, N) = \frac{n!}{(n-k)!k!}(e_T)^k(1-e_T)^{(N-k)}$$

dvs enligt en binomialfördelning. Skissa baserat på vad du kan hur man i princip, inga detaljer, med hjälp av Bayes sats och antagandet att $P(e_T, N)$ är konstant kan bestämma en fördelning för e_T (och därmed ett MAP estimat och/eller ett konfidensintervall för e_T).

Problem #2 (6p=3p+2p+1p)
a) Förklara följande begrepp: i) diskriminantfunktion, ii) beslutsgräns, iii) skillnaden mellan sannolikheten och sannolikhetstäthetsfunktionen för en stokastisk variabel som resulterar i observationer som betecknas x (eller x_n).
b) Vid klassning och regression med hjälp av kNN-metoden där k=3 finner man för en vektor $\mathbf{x}$ att de två närmaste prototypvektorerna tillhör klass 1, och att den tredje närmaste tillhör klass 2 av tre möjliga (klass1, klass 2 och klass3). Vidare

finner man att prototypvektorernas responsvärden är $y_1 = 1.03$, $y_2 = 0.96$ och $y_3 = 1.01$. Bestäm kNN-metodens klassificeringsbeslut och dess regressionsvärde i detta fall.
c) Man kan ju visa att vid Bayesiansk klassificering av två normalfördelade klasser som alla har samma kovariansmatris men olika medelväden så kan man reducera den optimala klassificeraren (som minimerar antalet felklassningar) till en enda linjär diskriminantfunktion. Antag att det finns N exempel tillgängliga för design och att medelvärden och kovariansmatris estimeras och att dessa estimat sedan används i uttrycket för den optimala linjära diskriminantfunktionen. Antag sedan att man också designar en linjär diskriminantfunktion med hjälp av minsta kvadratmetoden applicerad på de N designexemplen och att den resulterande linjära diskriminantfunktionen ger bättre resultat (i medeltal) för representativa testexempel än den diskriminantfunktion man får via Bayes beslutsteori för optimal klassning. Förklara hur en sådan situation kan uppstå! Obs! Antag att antalet testexempel är mycket stort så att möjligheten att den diskriminantfunktion som designats med hjälp av minsta kvadratmetoden bara ger ett sämre resultat på grund av tillfälligheter helt kan försummas.

Problem #3 (6p=3p+2p+1p)
a) Antag att en stokastisk variabel X har sannolikhetstäthetsfunktionen p(x). Skriv ned de integraler som definierar väntevärdet $m = E\{X\}$, variansen $\sigma^2 = E\{(X-m)^2\}$ och sannolikheten $P(X < m)$ att X är mindre än väntevärdet.
b) Antag att det finns N exempel tillgängliga för att estimera medelvärdet och kovariansmatrisen för en normalfördelning. Istället för att beräkna medelvärdet och kovariansmatrisen med hjälp av alla N exemplen kan man välja att INTE ta med de M stycken exempel som innehåller de M mest extrema (största och minsta) komponenterna. Vilken är den huvudsakliga fördelen respektive nackdelen med att göra så?
c) Antag att vektorn $\mathbf{x}$ är en realisering av en K-dimensionell normalfördelad stokastisk variabel med medelvärde noll och kovariansmatris C med egenvektorer $\mathbf{e}_k,\ \ k = 1, 2, \ldots, K$ och motsvarande egenvärden $\lambda_k,\ \ k = 1, 2, \ldots, K$. Hur stor är variansen längs riktningen som definieras av $\mathbf{e}_3$?

Problem #4 (6p=3p+2p+1p)
a) Förklara kort grundtankarna bakom PCA, hierarkisk klustring och Fishers diskriminant. Obs! Ge korta förklaringar (4-5 meningar) separat för en metod i taget. Gärna figurer!
b) Antag att alla mönster $\mathbf{x}_n$ i en datamängd har d dimensioner och längden ett, $||\mathbf{x}_n|| = 1$. Mönstren ligger alltså på en d-dimensionell sfär med radie ett. Antag att mönstren ligger i K tätt koncenterade kluster och att klustren är likformligt fördelade över sfären. Antag vidare att att det finns N exempel tillgängliga för

analys. Förklara varför resultatet av en principalkomponentanalys skulle indikera att ingen kompression skulle vara möjlig medan en klustringsanalys skulle indikera att varje observation kan komprimeras till ett enda heltal!
c) Antag att PCA utförs med hjälp av N stycken d-dimensionella mönster och att $d > N$. Bevisa eller förklara varför det i detta fall finns högst N stycken nollskilda (som inte är noll) egenvärden.

Problem #5 (6p=3p+2p+1p)
a) Förklara med hjälp av ett antal punkter (cirka 6 punkter) stegen i hur man med hjälp av PCA skulle kunna komprimera och sedan rekonstruera digitaliserade röntgenbilder på tex ett sjukhus. Förklara vad man behöver spara och vad man kan kasta bort, glöm inte bort medelvärdet! Inga formler krävs. Ge sedan också en formel som talar om hur många bilder man måste lagra för att faktiskt erhålla en minnesbesparing: Antag att varje bild består av N bildpunkter som reduceras till M rella tal och att varje reellt tal som lagras i datorn eller på hårddisken gör det i form av B bitar (dvs ett B bitar långt binärt tal). Obs! Ingen sönderklippning av de enskilda bilderna ska ske, varje röntgenbild/observation ska vara en kolumnvektor bestående av N element.
b) Vid komprimering av röntgenbilder på ett sjukhus, finns det någon praktisk vinst med att inte komprimera alla bilder med hjälp av samma egenvektorer utan att istället ha olika egenvektorer för olika bildgrupper? Glöm inte att motivera ditt svar, det är motiveringen som ger poäng. Förklara också kort hur man ska kombinera klustring och PCA för att kunna skapa en sådan uppdelning i grupper automatiskt.
c) Antag att PCA används för att lagra röntgenbilderna i ett sjukhus och att man behöver 500 egenvektorer för att få ett försumbart medel-rekonstruktionsfel. Antag vidare att det efter att dessa 500 egenvektorer identifierats och lagrats kommer in en ny typ av patienter till sjukhuset som innebär helt nya bilder. Dessa bilder är helt nya i den meningen att de INTE helt ligger i det underrum/hyperplan som spänns av de 500 redan identifierade egenvektorerna. Antag att egenvektorerna inte räknas om utan att de som redan räknats fram används direkt för kompression av de nya bilderna. Vad skulle detta få för praktiska konsekvenser för rekonstruktionsfelet och därmed också för risken att missa viktig information vid klinisk användning av de rekonstruktuerade bilderna?

Problem #6 (6p=3p+2p+1p)
a) Framförallt vid multivariat regression centrerar och normerar man ofta variablerna i ett första steg. Antag att alla ursprungsvariabler normerats så att de har medelvärde noll och varians ett och att en linjär modell $\tilde{y} = \tilde{\mathbf{a}}^T \tilde{\mathbf{x}}$ har skapats för de normaliserade variablerna $\tilde{y}$ och $\tilde{\mathbf{x}}$. Bestäm ett uttryck för motsvarande linjära modell i de ursprungliga variablerna.
b) Perceptronregeln är en felkorrigerande metod för linjära klassificerare. Vad menas med att den är felkorrigerande och varför är den inte lika praktiskt användbar som minsta kvadratmetoden?
c) Antag att en linjär klassificerare har $t_i = 1$ som target (önskat utvärde) då exempel kommer från klass ω_i och $t_i = 0$ då exempel kommer från någon annan klass. Notera att för ett givet mönster $\mathbf{x}$ gäller likheten $P(t_i = 1|\mathbf{x}) = P(\mathbf{x} \in \omega_i) = P(\omega_i|\mathbf{x})$ där $\mathbf{x} \in \omega_i$ betyder att $\mathbf{x}$ är dragen från klassen/fördelningen ω_i-Visa att (med detta val av target variabler) då gäller att $E\{t_i|\mathbf{x}\} = P(\omega_i|\mathbf{x})$ dvs att betingade väntevärdet för target-variabeln t_i för ett givet mönster är lika med sannolikheten att mönstret är draget från klass ω_i. Ledning: Använd definitionen av väntevärdet.

Lycka till!