Uppsala universitet
Signaler and System
MG 018 - 471 3229

Tentamen i datoriserad mönsterigenkänning

för Teknisk fysik E/Sy/B och Informationsteknologi,
Polacksbacken 9 januari 2004, 14.00-19.00

Hjälpmedel: Mathematical Handbook, Mathematics Handbook (BETA), räknare formelsamling på kurshemsidan.

Problem #1 (3p)
Civ ska bygga ett system för att klassificera handskrivna bokstäver som lästs in från en blankett och lagrats i form av en gråskalebild. Varje bild innehåller 700×500 bildpunkter och innehåller en handskriven bokstav. Varje bildpunkt innehåller ett gråskalevärde på intervallet $[0, 1]$.

a) Civ vill analysera strukturen i det bildarkiv hon har genom att göra dimensionsreduktion. Ge henne ett förslag på hur hon kan göra detta på ett lämpligt sätt med hjälp av en kombination av PCA och klustring.

b) Istället för att använda värdet i varje enkskild bildpunkt som ett individuellt särdrag, förklara varför Civ bör basera sitt system på en metod som klassificerar ett mindre antal särdrag som först extraherats från bilden.

c) Ge minst ett exempel på ett lämpligt särdrag som du tycker att Civ bör extrahera och motivera sedan ditt förslag. Ledning: rotations- och skalningsinvarians önskas

Problem #2 (3p)
Om en klassificerare testas med endast *ett* oberoende testexempel och klassificeraren misslyckades är det intressant att fråga sig hur mycket information vi fått om om klassificerarens sanna error rate e. Vi kan delvis svara på detta genom att för det allmänna fallet med k_t fel på N_t tester beräkna *a posteriori* fördelningen för e med hjälp av Bayes sats:

$$p(e|k_t, N_t) = Ke^k(1-e)^{N_t-k_t}p(e)$$

där
$$K = \frac{N_t!}{(N_t - k_t)!k_t!}$$
är en normaliseringskonstant och $p(e)$ är *a priori* fördelningen (täthetsfunktionen) för e.

a) Förklara varför valet $p(e) = 1$ är naturligt om vi inte har någon specialkunskap om problemet.

b) Bestäm och rita upp *a posteriori* fördelningen för e för fallet ovan dvs då man testat en klassificerare med ett exempel och exemplet blir felklassat.

c) Förklara vad det sk HPD-intervallet är för något. Påminnelse: HPD står för Highest Probability Density.

Problem #3 (3p)

a) Förklara grundtanken bakom "latent variable modeling" i samband med multivariat regression och exemplifiera med PCR.

b) Förklara varför PCR kan kasta bort värdefull information!

c) Förklara den grundläggande skillnaden mellan PCR och PLS.

Problem #4 (3p)
Vid linjär multivariat regression finns ett antal viktiga saker att tänka på.

a) Civ ska utföra ridge regression och måste därför ställa in värdet på regulariserings-parameten λ. Förklara hur Civ ska göra för att hitta det mest lämpliga värdet på λ.

b) Civ upptäcker att det bästa värdet på regulariserings-parametern λ är noll. Behövs ridge regression i detta fall? Hade Civ kunnat inse detta från början?

c) Visa att om man förbehandlar en variabel genom att först subtrahera bort medelvärdet och sedan dividera med variabelns standardavvikelse så får den nya variabeln medelvärde noll och varians ett.

Problem #5 (3p)
a) Beskriv ett praktiskt exempel där minsta antalet felklassningar *inte* är det naturliga valet. Glöm inte att motivera ditt svar.
b) Förklara begreppet risk i samband med Bayes beslutsteori och förklara vilket fall som ger beslut med minsta antalet felklassningar.
c) Förklara med hjälp av en figur iden bakom Fishers diskriminant.

Problem #6 (3p)
Antag att exempel från en klass (class ω_1) är dragna från en normalfördelining med medelvärde noll och varians ett och att data från en annan klass (klass ω_s) är dragna från en normalfördelning med medelvärde ett och varians två. **Obs! Både medelvärdena och variansera är olika!** Bestäm en optimal beslutsregel som minimerar antalet klassificeringsfel för data dragna från dessa fördelningar.

Problem #7 (3p)
Antag att det finns N stycken exempel på mönster med medelvärde noll i form av en mängd $D = \{\mathbf{x}_n\}, n = 1, 2, \ldots, N$. Kovariansmatrisen skattas som vanligt med

$$C = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_n \mathbf{x}_n^T = \frac{1}{N} X X^T$$

där $X = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$.
a) Visa att varje egenvektor $\mathbf{w}_p$ till C kan skrivas som en linjärkombination av exemplen, även om matrisen inte har full rang:

$$\mathbf{w}_p = \sum_{n=1}^{N} a_{np} \mathbf{x}_n = X \mathbf{a}_p$$

där $\mathbf{a}_p$ är en koefficientvektor.
b) Antag att vi har färre exempel än antalet dimensioner d, $N < d$ men att de är linjärt oberoende så att inversen till matrisen $X^T X$ existerar. Visa att $\mathbf{a}_p$ är egenvektor till matrisen $X^T X$. Obs! Notera att $X^T X \neq X X^T$.
c) Antag att $d = 1000$ och att det endast finns $N = 100$ linjärt oberoende exempel. Då har C endast rang 100 och alltså endast 100 egenvärden som inte är noll. Förklara varför resultatet i b) är intressant om man för detta exempel vill bestämma egenvektorerna till C. Ledning: Minne och beräkningstid.

Lycka till!