Uppsala universitet
Signaler and System
TO 018 - 471 3232

Tentamen i Mönsterigenkänning

Teknisk Fysik F3, Pollacksbacken, 10 januari 2001, 08.00-13.00

Hjälpmedel: Mathematics Handbook (BETA), miniräknare, formelsamling
OBS: Avsaknad av motiveringar ger poängavdrag.

Problem #1 (3+3=6p)
K-means clustering är ett typiskt exempel på en icke-lärarledd metod.
a) Beskriv tydligt denna algoritm (med formler och/eller figur där så krävs av tydlighetsskäl).
b) Anta att vi har funnit två kluster, C_1 och C_2, med centra i $\mathbf{m}_1$ respektive $\mathbf{m}_2$. Nya (oklassificerade) mönster tillkommer och vi klassificerar dessa m.h.a närmaste granne-metoden (vi använder standardkvadratnormen som avståndsmått). Detta förfarande i kombination med K-means clustering för att finna klustercentra har vissa likheter med Bayesiansk klassificering för normalfördelade data. Jämför ovannämnda förfarande (K-means+nearest neighbor) med ett förfarande där vi först skattar väntevärdesvektorer och kovariansmatriser för normalfördelningarna och sen applicerar Bayes beslutsteori. Vilka är likheterna och skillnaderna mellan dessa två olika förfaranden? Ange speciellt om det finns förutsättningar då de bör ge i princip samma resultat (med förutsättningar kan vi i detta fall räkna in även de antaganden vi själva gör).

Problem #2 (3+1+2=6p)
Betrakta problemet att approximera en 1-D funktion m.h.a en mycket enkel tvålagerperceptron som består av en neuron i vardera lagret
(dvs strukturen $x_i \rightharpoonup o \rightharpoonup o \rightarrow y_i$, med bias på de båda neuronerna). Till vår hjälp har vi en sekvens av träningsdatapar: $\{(x_1, d_1), (x_2, d_2), ..., (x_M, d_M)\}$ där indata, x_i, och börvärde, d_i, båda är skalärer.
a) Härled en träningsalgoritm för detta nät (baserad på gradientmetoden). Härledningen ska gälla s.k. *batch*-learning. Olinjäriteten är den vanliga logistiska funktionen, $\phi(v) = 1/(1 + e^{-v})$. OBS: Glöm ej bias.
b) Visa hur du ska modifiera algoritmen till att gälla s.k. pattern learning.
c) Ett problem med gradientmetoden är valet av steglängd. Ange och beskriv kortfattat två metoder för att hantera detta problem (Metoderna ska vara praktiska

i meningen att de inte tillför alltför mycket extra beräkningar i mer allmänna situationer då vi har ett stort antal vikter).

Problem #3 (2+2=4p)
Anta att vi ska komprimera tvådimensionella vektorer till en dimension m.h.a. principalkomponentanalys. Våra vektorer, $\mathbf{x}$, är dragna från en normalfördelning, $N(\mathbf{m}, \mathbf{C})$, med väntevärdesvektorn

$$\mathbf{m} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \text{och kovarianmatrisen} \quad \mathbf{C} = \begin{pmatrix} 3 & 1 \\ 1 & 4 \end{pmatrix}. \tag{1}$$

a) Informationen i den komprimerade vektorn används för rekonstruktion. Hur stort kommer rekonstruktionsfelet att vara i medeltal?
b) Utför kedjan av komprimering till 1-D med efterföljande rekonstruktion för vektorn $\mathbf{x}_i = [2, 2]^T$

Problem #4 (5p)
Betrakta ett mönsterigenkänningsproblem med två klasser, C_1 och C_2. Den tvådimensionella mönstervektorn, $\mathbf{x} = [x(1), x(2)]^T$, antas ha uppkommit genom en mätning som kan beskrivas som:

$$\begin{aligned} x(1) &= s_i(1) + n(1) \\ x(2) &= s_i(2) + n(2), \end{aligned} \tag{2}$$

för $i = 1$ eller $i = 2$. Vektorerna $\mathbf{s}_1 = [s_1(1), s_1(2)]^T$ och $\mathbf{s}_2 = [s_2(1), s_2(2)]^T$ antas vara kända. Termerna $n(1)$ och $n(2)$ är mätbrus som antas vara inbördes oberoende och oberoende av $\mathbf{s}_1$ och $\mathbf{s}_2$. Täthetsfördelningen för en komponent, $n(k)$, ges av

$$p\left(n(k)\right) = \frac{1}{K} e^{-|n(k)|/\lambda}, \tag{3}$$

där K är en normeringskonstant och parametern λ, som antas vara känd, definierar utsträckningen.
a) Härled den optimala beslutsregeln (minimalt antal felklassificeringar). A priorisannolikheterna för de båda klasserna är $P(C_1) = P_1$ respektive $P(C_2) = P_2$.
b) Visa i en figur hur beslutslinjen (alt. linjerna) går då $\mathbf{s}_1 = [1, 0]^T$ och $\mathbf{s}_2 = [-1, 0]^T$. Sätt $P_1 = e/(1+e)$ och $P_2 = 1/(1+e)$ samt $\lambda = 1/2$. Kommentera resultatet. Tillför mätningen av $x(2)$ något vid klassificeringen? Varför?

Problem #5 (2p)
Anta att en flerlagerperceptron skall utnyttjas för att klassificera mönster från klasserna $C_1, \ldots C_c$. Anta vidare att vi tränar nätet med ett mycket stort antal

träningsexempel och att nätets struktur har valts med tillräckligt rik struktur så att nätet kan approximera alla, för oss intressanta, funktioner godtyckligt nära. Under dessa förutsättningar kan man visa att utsignalen från nätet ges av det betingade väntevärdet:

$$\mathbf{y} = E[\mathbf{d}|\mathbf{x}]$$

där $\mathbf{x}$ är insignalen till nätet, $\mathbf{y}$ är utsignalsvektorn och $\mathbf{d}$ är börvärdesvektorn.

Anta nu att vi under träningen har valt $\mathbf{d}$ som

$$\mathbf{d} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}$$

där positionen för 1:an ges av klasstillhörigheten för det aktuella exemplet. Visa m.h.a. informationen som gavs ovan att signalen från utnod nr i för ett nytt mönster, $\mathbf{x}$, kommer att vara en god approximation av a posteriorisannolikheten $P(C_i|\mathbf{x})$.

Problem #6 (2p)
Beskriv tydligt hur korsvalidering kan användas för att finna en lämplig nätstruktur vid träning av en flerlagerperceptron.

Lycka till!