Uppsala universitet
Avd. för Signaler och System
TO 018 - 471 3082

Tentamen i Neuronnät och Lärande System

Teknisk Fysik F4, Pollacksbacken, 21 oktober 2002, 15.00-20.00

Hjälpmedel: Mathematics Handbook (BETA), miniräknare, formelsamlingar (deriveringsregler för matrisuttryck samt "Collection of Formulas for ...")
Motivera! Förtydliga ev. härledningar med hjälptext! Skriv tydligt!
Preliminär gräns för godkänt: 16p.

1. (4p) Täthetsskattning.
   Namnge och beskriv kortfattad fyra olika metoder som vi har tagit upp i kursen för att skatta täthetsfördelningar. Förklara dessutom, där det är möjligt, vilka underliggande antaganden som ligger till grund för metoderna.

2. (2+2+1+2+1=8p) Anta att variablerna t och x är beroende via sambandet

$$t(x) = h(x) + e, \tag{1}$$

   där $h(x)$ är en okänd funktion och e är en brusterm som är okorrelerad med x och som har väntevärde 0 och varians σ^2. Vi vill approximera $h(x)$ mha funktionen $g(x, \hat{\mathbf{w}})$, där $\hat{\mathbf{w}}$ är en parametervektor som vi har skattat med hjälp av en träningsmängd, $\{(x_1, t_1), ..., (x_N, t_N)\}$. Anta nu att vi vill analysera hur stort medelapproximationsfel vi kommer att ha (medelvärdesbildning över alla tänkbara träningsmängder, och över alla x). Vi är intresserade att erhålla en approximation, $g(x, \hat{\mathbf{w}})$, som minimerar det s.k. generaliseringsfelet, $J_{gen} = E[(t - g(x, \hat{\mathbf{w}}))^2]$. Vi kan dela upp J_{gen} som

$$J_{gen} = E[(< g(x, \hat{\mathbf{w}}) > -h(x))^2] + E[(g(x, \hat{\mathbf{w}}) - < g(x, \hat{\mathbf{w}}) >)^2] + \sigma^2 \tag{2}$$

   (a) Härled sambandet i ekv. (2).
   (b) Vad brukar de två första termerna i uppdelningen kallas? Förklara dessutom i ord vad de står för.
   (c) Under lyckosamma omständigheter kan vi få de två första termerna att gå mot noll så uppenbarligen är σ^2 det minsta möjliga värdet som J_{gen} kan anta. Ge en enkel förklaring till varför detta måste vara den undre gränsen.
   (d) Validering och korsvalidering kan användas för bestämning av en lämplig modellstruktur. Beskriv hur de fungerar. Peka ut skillnaderna.
   (e) Motivera mha uttrycket i ekv. (2) varför validering (någon av varianterna) bör resultera i en modell med goda generaliseringsegenskaper.

3. (2+2=4p) Anta att du har tränat en flerlagerperceptron (MLP) för att lösa ett mönsterigenkänningsproblem med L st. klasser. Du har under träningen haft tillgång till en mycket stor mängd träningsdata i alla områden av mönsterrummet och vid träningen har du minimerat det vanliga kvadratfelskriteriet. Nätet antas ha många vikter och därmed goda möjligheter att, med god noggrannhet, approximera de flesta funktioner av intresse.

(a) Anta att du väljer ett nät med L st. utneuroner. Hur ska du välja dina börvärdesvektorer vid träning, $\mathbf{t}_1, ..., \mathbf{t}_N$, så att du som utsignal från neuron nr m för ett nytt mönster $\mathbf{x}$ får en approximation av *a posteriori* sannolikheten $P(C_m|\mathbf{x})$? (Ingen härledning efterfrågas)

(b) Betrakta nu ett problem med två klasser (dvs $L = 2$). Anta att *a priori* sannolikheterna för de två klasserna för träningsdata var $P(C_1) = P$ respektive $P(C_2) = 1 - P$. Anta vidare att det kommer till din kännedom att förutsättningarna har ändrats efter träningen så att *a priori* sannolikheterna numera är $P(C_1) = P'$ samt $P(C_2) = 1 - P'$. Visa hur du kan, utan att träna om nätet, utnyttja denna nya information till att korrigera skattningar av *a posteriori* sannolikheterna $P(C_1|\mathbf{x})$ resp. $P(C_2|\mathbf{x})$ som ditt nät producerar på sina två utgångar. (Ledning: Betrakta kvoten mellan $P(C_1|\mathbf{x})$ och $P(C_2|\mathbf{x})$ och utnyttja Bayes sats samt summaregeln för uteslutande händelser.)

4. (2+1=3p) Självorganiserande kartor (SOM).

(a) Beskriv mha en figur strukturen och funktionen (ej träning) hos en självorganiserande karta.

(b) Förklara i detalj vad som menas med att nätet är topologiskt ordnat efter träningen.

5. (3+2+2=7p) RBF och MLP.

(a) Beskriv mha figurer och formler uppbyggnaden av en radiell basfunktion (RBF) repektive en flerlagerperceptron (MLP). Välj strukturen på MLP:n så att den lämpar sig för att lösa allmänna funktionsapproxmationsproblem.

(b) Träning av RBF brukar delas in i två faser, vilka? Beskriv kortfattat vad som sker i dessa faser.

(c) Härled ett slutet uttryck för parameterinställningen i den andra träningsfasen.

6. (2+2=4p) Gömda Markovmodeller

(a) Förklara i detalj hur en s.k. standard HMM är uppbyggd. Speciellt, förklara vilken funktion s.k. *insert states* har. Utgå från problemet att känna igen textsträngar när du beskriver modellen och förklarar de ingående parametrarna.

(b) Bakåtalgoritmen är en rekursion för att beräkna $\beta_i(t) = P(X^1, ..., X^t, s^t = q_i|w)$, dvs sannolikheten för att den observerade sekvensen upp till tiden t är $X^1, ..., X^t$, samt att tillståndet vid tiden t är q_i (givet en modell med kända parametrar, w). Anta att vi vill bestämma det mest sannolika tillståndet vid tiden t givet en observation, $O = \{X^1, ..., X^T\}$. Visa varför $\beta_i(t)$ behövs för göra detta.

Lycka till!