Uppsala universitet
Signaler and System
TO 018 - 471 3082

Tentamen i Neuronnät och Lärande System

Teknisk Fysik F4, Pollacksbacken, 20 oktober 2003, 08.00-13.00

Hjälpmedel: Mathematics Handbook (BETA), miniräknare, formelsamling (deriveringsregler för matrisuttryck)
Motivera! Förtydliga ev. härledningar med hjälptext! Skriv tydligt!
Preliminär gräns för godkänt: 15p.

1. (1+1+2+1+1=6p) Blandat.
   (a) I vilket sammanhang förekommer “leave-one-out”? Ge en kort redogörelse.
   (b) Vad är en Gauss-mixtur?
   (c) Vad menas med regularisering och vilket problem försöker man lösa med denna teknik? På vilket sätt kan man säga att regularisering är relaterat till maximum a posteriori (MAP) skattning?
   (d) I samband med klassificering brukar begreppet indikatorfunktion förkomma. Hur definieras indikatorfunktionen, $t(\mathbf{x})$?
   (e) Vad är det betingade väntevärdet av indikatorfunktionen, dvs $E[t(\mathbf{x})]$ (man brukar kanske oftare skriva $E[t|\mathbf{x}]$, vilket betyder samma sak.)

2. (3+1=4p) RBF.
   (a) Träning av ett nät av radiella basfunktioner sker oftast i två fristående faser: (i) inställning av centrum för basfunktionerna, $\mathbf{m}_1, ..., \mathbf{m}_M$, och (ii) inställning av vikterna i nätet, $\mathbf{w} = (w_1, ..., w_M)^T$. Härled, genom att ställa upp och minimera en kriteriefunktion, ett uttryck för dessa vikter. Utgå från följande variabler:

$$\bar{\phi}(\mathbf{x}_n) = \begin{pmatrix} \phi(\mathbf{x}_n - \mathbf{m}_1) \\ \vdots \\ \phi(\mathbf{x}_n - \mathbf{m}_M) \end{pmatrix} \quad , \quad \Phi = \left(\bar{\phi}(\mathbf{x}_1)\ \bar{\phi}(\mathbf{x}_2)\ \ldots \bar{\phi}(\mathbf{x}_N)\right) \quad \text{och} \quad \mathbf{T} = (t_1\ t_2\ ...t_N) \tag{1}$$

   där M är antalet basfunktioner och N är antalet träningsdata. Som vanligt är $\mathbf{x}$ indata och t börvärde.
   (b) Vilken är den väsentliga fördelen med att dela upp träningen i två faser?

3. (1+1+1+2+2=7p) Självorganiserande nät (SOM)
   (a) Vad står förkortningen BMU för?
   (b) Vad menas med att en tränad SOM är *topologiskt ordnad*? Ge definitionen!
   (c) Vad menas med den s.k. *grannskapsfunktionen* (eng. neighborhood function)?.
   (d) Beskriv träningen av en SOM. För full poäng, skriv ner träningsalgoritmen och förklara speciellt vilka roller BMU och grannskapsfunktionen spelar under träningen.

(e) Träning av SOM uppvisar stora likheter med KMC algoritmen. Om vi antar att vi har obegränsat med data från en fördelning med PDF $p(\mathbf{x})$ så kan KMC tolkas som att man söker minimera en kriteriefunktion (distortionsmått) som har formen $D = \int ||\mathbf{x} - \mathbf{m}(c(\mathbf{x}))||^2 p(\mathbf{x}) d\mathbf{x}$, där $\mathbf{m}$ är en prototypvektor med index $c(\mathbf{x})$ (c står för "code"). Träningsalgoritmen för SOM kan på ett liknande sätt förklaras som minimering av ett annat distortionsmått. Vilket? Ge en motivering till detta distortionsmått.

4. (7p) HMM. Beskriv i detalj hur du med hjälp av HMM:er kan bygga upp ett system för taligenkänning där mängden av ord är begränsade till ensiffriga tal (0-9). Ge din lösning som en projektbeskrivning där du (i) kort beskriver vilka träningsdata du förutsätter ska finnas, (ii) anger och motiverar hur dessa ska förprocessas, (iii) anger och motiverar vilka algoritmer du tänker använda för träning, (iv) hur du bestämmer lämpliga strukturer hos dina HMM:er, speciellt hur du kan tänkas utnyttja träningsdata även för denna uppgift, samt (v) redogör för hur du slutligen ska använda dina HMM:er för klassificering av nya talade siffror.

5. (2+1+2+1=6p) MLP

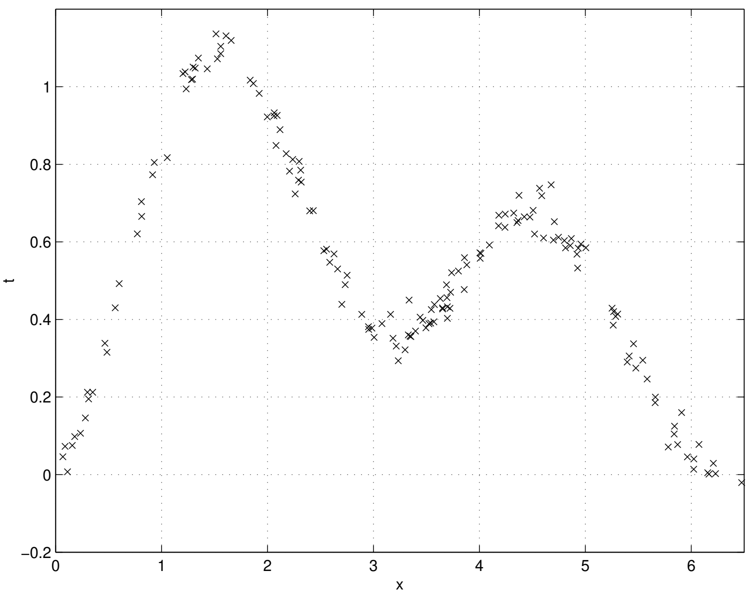
(a) Betrakta träningsdata i Figur 1 på nästa sida. Rita i detalj upp en MLP som bör klara att med god noggrannhet approximera t givet x.

(b) Ange dessutom lämpliga värden på vikter etc. i ovanstående nät så att denna approximation blir god. (Gör inställningen "för hand").

(c) Vid träningen av en MLP försöker man som bekant minimera en kriteriefunktion. Trots att Newtons metod för optimering är känd för att ha mycket goda konvergensegenskaper så används många gånger i praktiken istället någon metod som utnyttjar endast gradientinformation. Ange två tänkbara skäl till detta.

(d) Vad heter metoden att beräkna gradienter i samband med MLP:er? Motivera kortfattat detta namn.

Lycka till!



Figur 1: Träningsdata i uppgift 5