Uppsala universitet
Signaler and System
TO 018 - 471 3232

Tentamen i Mönsterigenkänning

Teknisk Fysik F3, Pollacksbacken, 15 december 2000, 14.00-20.00

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

Problem #1 (2+1+1=4p)
a) Beskriv m.h.a. ett flödesschema hur K-närmaste granne algoritmen fungerar.
b) Den kanske mest uppenbara tillämpningen av denna algoritm är *klassificering*. Den kan dock användas också för *estimering*. Beskriv hur.
c) Bortsett från problem som är relaterade till beräkningstyngd etc., ange ett fundamentalt problem vid användandet av närmaste granne-metoder. Ledn: Betrakta ett exempel med mönstervektorer där komponenterna beskriver vitt skilda särdrag (ex. pH-värde, temperatur etc.).

Problem #2 (4p)
Anta att en mätutrustning kontinuerligt mäter upp data som lagras i en serie vektorer, $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k, ...$. Informationen i dessa vektorer skall sändas över en kanal med alltför låg kapacitet och någon form av datakompression är därför önskvärd. Visa i detalj hur en sådan datakompression kan utföras m.h.a. principalkomponentanalys (PCA) samt hur approximationer av **x**-vektorerna kan återskapas på mottagarsidan. (Ange tydligt vilka data som till slut skickas över kanalen, hur dessa data beräknas, hur data återskapas, hur de variabler som ingår i beräkningarna erhålls)

Problem #3 (2+2+1=5p)
Ett återkommande problem vid användandet av en flerlagerperceptron (MLP) är s.k. överanpassning.
a) Utnyttja ett exempel med funktionsapproximation för att beskriva problemet i detalj.
b) Beskriv en metod som kan användas för att detektera (med rel. stor sannolikhet) om du står inför överanpassningsproblemet vid en viss tillämpning.
c) Beskriv en iterativ metod som utnyttjar metoden i b) för att undvika problem med överanpassning.

Problem #4 (2+2=4p)
Betrakta ett mönsterigenkänningsproblem med två klasser, C_1 och C_2. Mönstervek-

torn, $\mathbf{x}$, antas vara normalfördelad med väntevärde $\mathbf{m}_1$ och kovariansmatris Σ_1 för klass C_i samt $\mathbf{m}_2$ och Σ_2 för klass C_2. Vidare antas *a priori* sannolikheterna vara $P(C_1) = P_1$ samt $P(C_2) = P_2$.
a) Härled ekvationen för de optimala beslutsytorna (i betydelsen minst antal förväntade felklassificeringar). Ekvationen skall skrivas på formen $\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c = 0$.
b) Under vilka förutsättningar är beslutsytorna linjära för detta mönsterigenkänningsproblem?

Problem #5 (1+1=2p)
För ett givet mönsterigenkänningsproblem har man bland annat att välja mellan modellbaserade (parametriska) metoder eller icke-parametriska (i boken "data-driven") metoder.
a) Bayes beslutsteori baserat på tex Gaussiska fördelningar är ett exempel på en parametrisk metod. Vad definierar modellen i detta fall?
b) Ange under vilka förutsättningar valet rimligtvis bör/måste falla på en icke-parametrisk metod. Du ska i ditt svar kunna argumentera mot en sunt skeptisk person som frågar varför du inte använder tex Bayes beslutsteori (som ju faktiskt ska ge *optimala* beslutsregler).

Problem #6 (2p)
Förklara principen för hur ett binärt beslutsträd fungerar och kan användas.

Problem #7 (1+2+1=4p)
Logistisk regression.
Anta att du har tillgång till en serie träningsdata, $\{(\mathbf{x}_1, d_1), (\mathbf{x}_2, d_2), ..., (\mathbf{x}_P, d_P)\}$, där d_k är det skalära börvärdet till den N-dimensionella vektorn $\mathbf{x}_k$. Uppgiften är att m.h.a träningsdata finna approximationer av d_k med $y_p = \phi(\mathbf{w}^T\mathbf{x})$, där $\phi(z) = \frac{1}{1+e^{-z}}$.
a) Ställ upp det ofta förekommande minsta kvadrat-kriteriet som används för att ta fram den optimala vektorn $\mathbf{w} = (w_1, ..., w_N)^T$.
b) Optimeringen av $\mathbf{w}$ sker oftast iterativt i form av en gradientsökning. Härled en uppdateringsregel för vikten w_i på formen $w_i(n+1) = w_i(n) + \eta g(\cdot)$ där $g(\cdot)$ är en funktion av vikterna och dina träningsdata. (n är iterationsindex)
c) Man brukar göra skillnad på s.k. batch learning och pattern learning. Vari består skillnaden och hur ska uppdateringsregeln i uppgift b) modifieras för att erhålla pattern learning-varianten?

Lycka till!