Uppsala universitet
Signaler and System
MG 018 - 471 3229

Tentamen i mönsterigenkänning

Teknisk fysik, 3E/Sy 10 januari, 1998, Polacksbacken 8.00-14.00

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

Problem #1 (2+1=3p)
Som expert har du blivit anlitad för att analysera spektrala data i en livsmedelstillämpning. Uppgiften består i att med hjälp av uppmätta absorbtionsspektra för 10000 frekvenser i våglängdsområdet 600-1000 nm kunna förutsäga halten av protein, fett och vatten i ett prov. Till ditt förfogande finns 200 exempel som dokumenterats med hjälp av analytisk kemi. Vidare finns en nästan obegränsad tillgång på odokumenterade data on-line i den livsmedelsfabrik där systemet är tänkt att fungera.
a) Förklara i ett antal steg hur man med hjälp av principalkomonentanalys kan reducera antalet dimensioner.
b) Efter dimensionsreduktionen sker en klassning av innehållet i varje prov med hjälp av den reducerade särdragsvektorn. Förklara *med hjälp av en figur* varför det kan vara farligt att reducera dimensionen på indata då egenvärdena på de trunkerade komponenterna inte är identiskt lika med noll.

Problem #2 (2+1=3p)
Ett alternativ till linjära diskriminantfunktioner är kvadratiska diskriminantfunktioner, $g_i(\mathbf{x}) = \mathbf{x}^T Q_i \mathbf{x} + \mathbf{q}^T \mathbf{x} + q_i 0$.
a) För tvåklassproblem, förklara matematiskt skillnaden i form mellan de beslutsytor som en klassificerare som arbetar med linjära respektive kvadratiska diskriminantfunktioner kan skapa.
b) Under vilket villkor på Q_i blir den klassificerare som är baserad på kvadratiska diskriminantfunktioner ekvivalent med en linjär klassificerare?

Problem #3 (1+1=2p)
a) Förklara vad en outlier är och varför sådana är farliga vid skattning av parametrar som medel och kovariansmatris.

b) Om man antar att data genereras av en flerdimensionell normalfördelning när data egentligen genereras av en gaussian mixture stämmer inte modell och verklighet. Ge ett eget exempel på en situation där ett modellfel av denna typ gör att den klassificerare man får under antagande om gaussiska fördelningar kan bli helt oanvändbar.

Problem #3 (3p)
a) Härled error backpropagation för ett mycket litet neuronnät betstående av *två* inputs, *en* gömd neuron och *en* utneuron. Du behöver inte ta med bias men noderna måste vara olinjära.

Problem #5 (2+2=4p)
a) Antag att de enda möjliga binära mönster som genereras från ett system är $\mathbf{x}_1 = [1,1,0]^T$, $\mathbf{x}_2 = [1,0,1]^T$,$\mathbf{x}_3 = [0,1,1]^T$ and $\mathbf{x}_1 = [0,0,0]^T$, alla med sannolikheten p=1/4. Uppenbarligen finns har en viss struktur eftersom fyra möjliga mönster som t.ex. $[1,1,1]^T$ inte genereras. Beräkna väntevärde och kovariansmatris för fördelningen.
b) Visa att för en fördelning där alla tänkbara mönster genereras, var och en med sannolikheten p=1/8, ger samma medel och kovariansmatris som för fördelningen i deluppgift a).

Problem #6 (2+2=4p)
Förklara hur man med hjälp av PCA kan komprimera 1000 bilder genom att
a) Dela in bilderna i små delbilder.
b) Arbeta med hela bilder.

Problem #7 (3p)
Dimensionsreduktion är en viktig komponent i många system för mönsterigenkänning. De två klassiska metoderna är PCA och klustring. Respektive metod bygger på speciella antaganden om hur mönstren fördelar sig i mönsterrummet. Förklara *med hjälp av figurer* vilket grundantagande som gäller för respektive metod.

Problem #8 (3p)
a) Förklara hur en flerlagerperceptron kan användas som en olinjär diskriminantfunktion.
b) Förklara begreppet korsvalidering i samband med träning av flerlagerperceptroner.
c) Förklara varför den klassiska gradientmetoden kallas för error backpropagation då den appliceras på flerlagerperceptroner.

Lycka till!