Neurovõrgud. Praktikum 7.

1. aprill 2005. a.

1 Bayesi klassifitseerija

Klassifitseerimine on teatud objekti liigitamine ühte ettantud klassi. Klassifitseerija all mõistame funktsiooni g, mis saab ette objekti ,,kirjelduse" tema teatud *tunnuste* abil, ning väljastab arvu k, mis tähistab selle objekti klassi (mis ei pruugi olema alati õige). Siinkohal vaatleme ainult juhtu kus objektide tunnused on reaalarvulised vektorid, ning klasside arv on 2. Sel juhul on klassifitseerija funktsioon

$$g : \mathbb{R}^d \to \{1, 2\}$$

Näiteks võiks otsida klassifitseerijat, mis inimese kasvu ja kaalu järgi ütleks, kas see on mees või naine.
Selge et tihti leiduvad ühe ja sama tunnusvektoriga objekte, mis kuuluvad erinevatesse klassidesse. Kuna klassifitserija peab tunnusvektori järgi alati kindlalt otsustada ühte klassi, siis mõnede objektide puhul võib ta eksida. Meid huvitab selline klassifitseerija, mille puhul eksimise tõenäosus on minimaalne.
Tähistame tunnusvektoreid $\mathbf{x}$-ga, ning klasse y-ga. $F(\mathbf{x}, y)$ olgu vektorite ja klasside ühisjaotus. Klassifitseerija g *riskiks* $R(g)$ nimetame tõenäosust, et jaotusest F juhuslikult valitud objekt on valesti klassifitseeritud g poolt:

$$R(g) = P\{g(\mathbf{x}) \neq y\},\ (\mathbf{x}, y) \leftarrow F$$

Seda on mugav panna veidike teisel kujul. Defineerime *kaofunktsiooni* $L(y_1, y_2)$

järgmiselt: $L(y_1, y_2) = 1$ siis kui $y_1 \neq y_2$ ja 0 vastasel juhul. Siis[1]

$$R(g) = \int L(y, g(\mathbf{x}))dF(\mathbf{x}, y)$$

Üldjuhul võib kaofunktsioon olla ka teistsugune, siis ei ole risk enam vea tõenäosus, vaid n.ö. *keskmine kahju*. Näiteks kui meie jaoks oleks klassi 1 objektide äratundmine kaks korda olulisem kui klassi 2, võiksime me määrama $L(1, 2) = 2$. Siinkohal me sellist juhtumit ei vaatle ning tegeleme ainult eespool kirjeldatud *sümmeetrilise* kaofunktsiooniga.
Hea klassifitseerija on siis selline, mille puhul risk (antud juhul vea tõenäosus) on väike. Fikseerime nüüd mingi konkreetne tunnusvektor $\mathbf{x}_0$. Oletame et klassifitseerija g seob talle vastavusse klassi k. Sel juhul eksib ta tõenäosusega $P\{y \neq k \,|\, \mathbf{x}_0\}$. Seda tõenäosust nimetatakse *tinglikuks riskiks*, ning tähistatakse $R(g \,|\, \mathbf{x}_0)$:

$$R(g \,|\, \mathbf{x}_0) = \int L(y, g(\mathbf{x}_0))dF(y \,|\, \mathbf{x}_0) =$$

$$= L(1, g(\mathbf{x}_0))P(1 \,|\, \mathbf{x}_0) + L(2, g(\mathbf{x}_0))P(2 \,|\, \mathbf{x}_0) = \begin{cases} P(1 \,|\, \mathbf{x}_0), & \text{kui } g(\mathbf{x}_0) = 2 \\ P(2 \,|\, \mathbf{x}_0), & \text{kui } g(\mathbf{x}_0) = 1 \end{cases}$$

Mõistlik on valida vektori $\mathbf{x}_0$ klassiks seda klassi, mille puhul vea tõenäosus (ehk tinglik risk) on väiksem. Osutub, et klassifitseerija, mis iga $\mathbf{x}$ korral minimiseerib tingliku riski, minimiseerib ka päris riski $R(g)$. Sellist klassifitseerijat nimetatakse *Bayesi klassifitseerijaks.*

Ülesanne 1 (1p): Näita et klassifitseerija, mis iga $\mathbf{x}$ korral minimiseerib tingliku riski, minimiseerib ka päris riski $R(g)$.
Vihje: näita et $R(g) = \int R(g \,|\, \mathbf{x}_0)dF(\mathbf{x}_0)$.

Ülesanne 2 (1p): Praktikas kirjutatakse Bayesi klassifitseerimisreeglit kujul:

> Kui $P(\mathbf{x}_0 \,|\, 1)P(1) > P(\mathbf{x}_0 \,|\, 2)P(2)$, siis klassifitseeri objekti $\mathbf{x}_0$ klassi 1.
> Vastupidisel juhul — klassi 2.

Näita et see on samaväärne algse sõnastusega. Vihje: kasuta Bayesi reeglit.

[1] See on nn. *Lebesque-Stieltjes*-i integral üle jaotuse. Kui jaotusel F leidub tihedusfunktsioon $f(\mathbf{x}, y)$ siis suvalise h korral

$$\int h(\mathbf{x}, y)dF(\mathbf{x}, y) = \int h(\mathbf{x}, y)f(\mathbf{x}, y)d\mathbf{x}dy$$

Kui aga F on diskreetne jaotus objektide hulgal $\mathbf{x}_0, \mathbf{x}_1, \ldots$, siis avaldub integral summana:

$$\int h(\mathbf{x})dF(\mathbf{x}) = \sum_{i=0}^{\infty} h(\mathbf{x}_i)P\{X = \mathbf{x}_i\})$$

Ülesanne 3 (2p): Oletame et tudengi õppeedukus Y sõltub kolmest faktorist: I — kui palju tunde päevas tudeng viibib internetis, B — kui palju liitrit õlut ta nädalas joob, E — teatud tudengi laiskuse hinnang. Olgu need suurused paarikaupa sõltumatud, ning eksponentsiaalse jaotusega: kõigi nende suuruste tihedusfunktsiooniks olgu e^{-x}.([2])
Oletame et õppeedukus on määratud lihtsa mudeliga: kui $I + B + E \leq 7$, siis tudeng õpib hästi ($Y = 1$), vastasel juhul — halvasti ($Y = 2$). Seega kui me teaksime antud tudengi kohta need kolm väärtust, saaksime me kohe õelda kui hea ta on. Tegelikkuses teame me ainult suurusi I ja B, ning soovime nende suuruste järgi kõige kindlamalt otsustada Y väärtust. Leia Bayesi klassifitseerijat sellel juhul. Leia tema riski. Milline oleks konstantse klassifitseerija $g(I, B) = 1$ risk?
Vihjed:

1. Leia suurust $P\{I + B + E \leq 7 \,|\, I, B\}$
2. Uuri millal on see võrdne $\frac{1}{2}$. Seal on klassifitseerijal eraldav joon.
3. Riski leidmiseks integreeri vea tõenäosus kahes piirkonnas eraldi.

[2] Seega $P\{I \leq x\} = 1 - e^{-x}$.

2 Bayesi klassifitseerija normaalse jaotusega klasside eristamiseks

Vaatleme kahemõõtmelisi tunnusvektoreid kahest klassist. Olgu mõlema klassi vektorid normaaljaotusega, täpsemalt

$$P(\mathbf{x}|i) = \frac{1}{2\pi\sqrt{\det(\mathbf{\Sigma}_i)}} \exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu}_i)^T \mathbf{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)}{2}\right)$$

kus $\boldsymbol{\mu}_i$ ja $\mathbf{\Sigma}_i$ on klassi i jaotuse keskmine ja kovariatsiooni maatriks, ning

$$\boldsymbol{\mu}_1 = (0,0)^T, \quad \mathbf{\Sigma}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{\mu}_2 = (1,3)^T, \quad \mathbf{\Sigma}_2 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

Klassid olgu võrdse tõenäosusega: $P(1) = P(2) = 0.5$

Ülesanne 4 (1p): Genereeri igast klassist 100-1000 esindajat, pane need punktidena tasandile. Konstrueeri Bayesi klassifitseerija antud klasside jaoks. Konstrueeri samale graafikule klasse eraldavat joont. Tee seda veel kord parameetritega

$$\boldsymbol{\mu}_0 = (0,0)^T, \quad \mathbf{\Sigma}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0.25 \end{pmatrix}, \quad \boldsymbol{\mu}_1 = (0,2)^T, \quad \mathbf{\Sigma}_1 = \begin{pmatrix} 0.25 & 0 \\ 0 & 1 \end{pmatrix}$$

Esita kood ja saadud pildid.
Vihjed:

1. Juhuslike punktide genereerimiseks võib kasutada järgmist koodi:

```
m1 = [0 0]';
cov1 = diag([1 1]);
X = grand(300, 'mn', m1, cov1); // mn = 'multivariate normal'
```

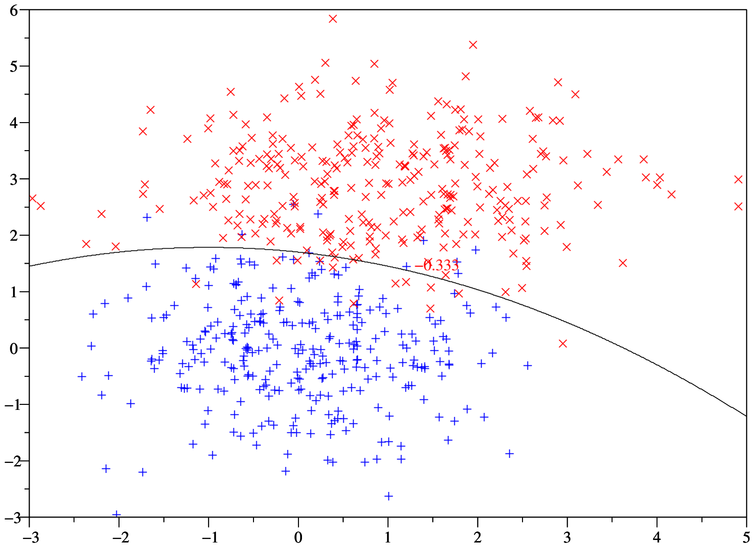
2. Klasside eraldava joone konstrueerida saab näiteks niimoodi[3]:

```
t=[-4:0.1:7];
fcontour2d(t,t,g,2,style=1);
// Siin g on klassifitseerija ise, s.t.
//  funktsioon, mis saab ette (x,y), ning
//  tagastab klassi: 1 voi 2. "function y=g(x1,x2)"
```

3. Vaja läheb ka sellist funktsiooni (normaaljaotuse tihedust):

```
function p = gaussian(x, y, m, cov)
  p = 1/(2*%pi*sqrt(det(cov))) * ...
      exp(-1/2*(([x;y]-m)'*inv(cov)*([x;y]-m)));
endfunction
```

[3] Veel üks võimalus oleks kasutada funktsiooni `fgrayplot`



Joonis 1: Bayes'i klassifitseerija normaalselt jaotatud klasside jaoks

Ülesanne 5 (1p): Hinda konstrueeritud klassifitseerija riski (empiiriliselt). Esita kood ja saadud näitaja.

Ülesanne 6 (2p): Lihtne näha et gaussi klasside juhul on Bayesi klassifitseerija eraldav joon alati ülimalt teise astme kõver. Osutub et kui klasside jaotuste kovariatsioonimaatriksid on võrdsed, keskväärtused erinevad ja klassid on võrdtõenäosed, on bayesi klassifitseerija eraldavaks jooneks alati sirgjoon. Tõesta seda.

Ülesanne 7 (1p): Bayesi klassifitseerija konstrueerimiseks on vaja teada tunnusvektorite jaotust, mida praktikas teada ei saa. Kas siis on see kõik ainult teoreetiline konstruktsioon või saab seda ka päriselus rakendada? Miks/Kuidas?