Neurovõrgud. Praktikum 7.

Ülesannete lahendused

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

Lahendus: Kuna

$$R(g) = \int_{X \times Y} L(y, g(\mathbf{x})) dF(\mathbf{x}, y) = \int_X \left(\int_Y L(y, g(\mathbf{x})) dF(y \,|\, \mathbf{x}) \right) dF(\mathbf{x}) =$$
$$= \int_X R(g \,|\, \mathbf{x}) dF(\mathbf{x})$$

siis minimiseerides iga $\mathbf{x}$ korral tingliku riski $R(g \,|\, \mathbf{x})$ minimiseeritakse ka kogurisk.

Ülesanne 2: 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.

Lahendus: Jagame mõlemad võrratuse pooled $P(\mathbf{x}_0)$-ga läbi ning kasutame Bayesi valemit:

$$\frac{P(\mathbf{x}_0 \,|\, 1)P(1)}{P(\mathbf{x}_0)} = P(1 \,|\, \mathbf{x}_0)$$

millega viime reeglit kohe originaalkujule („kui $P(1 \,|\, \mathbf{x}_0) > P(2 \,|\, \mathbf{x}_0)$ siis klass 1, vastasel juhul 2").
$P(\mathbf{x}_0)$-ga jagamine on lubatud sest klassifitseeritava $\mathbf{x}_0$ esinemise tõenäosus (või tõenäosuse tihedus pideva muutuja korral) on alati nullist erinev.

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}.
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 öelda 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?

Lahendus: Klasside a-posteriori tõenäosused on

$$\begin{aligned} P(Y = 1 \,|\, I = i, B = b) &= P(I + B + E \leq 7 \,|\, I = i, B = b) = \\ &= P(E \leq 7 - i - b) = 1 - e^{-(7-i-b)} \\ P(Y = 2 \,|\, I = i, B = b) &= 1 - P(Y = 2 \,|\, I = i, B = b) = e^{-(7-i-b)} \end{aligned}$$

Tõenäosused on võrdsed siis kui $e^{-(7-i-b)} = \frac{1}{2}$, s. t. kui $i+b = 7-\ln(2)$. Seega Bayesi klassifitseerija g^* on siin lineaarne: kui $i + b \leq 7 - \ln(2)$ klassifitseerib ta tudengit klassi 1 (hea), vastasel juhul — klassi 2.
Leiame nüüd tema riski (meie kontekstis siis vea tõenäosuse):

$$\begin{aligned} R^* = R(g^*) = \mathbb{E}_{I,B,Y}\{L(Y, g^*(I, B))\} &= \int_{I\times B\times Y} L(Y, g^*(I, B)) dF(I, B, Y) = \\ = \int_{I+B\leq 7-\ln(2)} L(Y, 1) dF(I, B, Y) &+ \int_{I+B>7-\ln(2)} L(Y, 2) dF(I, B, Y) \end{aligned}$$

Vaatleme esimest liidetavat:

$$\begin{aligned} \int L(Y, 1) dF(I, B, Y) &= \int_{I\times B} \left(\int_Y L(Y, 1) dF(Y \,|\, I, B) \right) dF(I, B) = \\ &= \int_{I\times B} (L(1, 1) P(Y = 1 \,|\, I, B) + L(2, 1) P(Y = 2 \,|\, I, B)) dF(I, B) = \\ &= \int P(Y = 2 \,|\, I, B) dF(I, B) = \int P(E + I + B > 7 \,|\, I, B) dF(I, B) = \\ &= \int P(E > 7 - I - B \,|\, I, B) dF(I, B) = \int e^{-(7-I-B)} dF(I, B) = \\ &= \int_{i+b\leq 7-\ln(2)} e^{-(7-i-b)} e^{-i} e^{-b} di\, db = \\ &= \int_{i+b\leq 7-\ln(2)} e^{-7} di\, db = \frac{1}{2} e^{-7} (7 - \ln(2))^2 \approx 0.018 \end{aligned}$$

analoogselt teise liidetavaga:

$$\begin{aligned}
&\int L(Y,2)dF(I,B,Y) = \int_{I+B>7-\ln(2)} P(Y=1\,|\,I,B)dF(I,B) = \\
&= \int P(E+I+B \le 7\,|\,I,B)dF(I,B) = \\
&= \int P(E \le 7-I-B\,|\,I,B)dF(I,B) = \\
&= \int \max((1-e^{-(7-I-B)}),0)dF(I,B) = \\
&= \int_{7-\ln(2)<i+b\le 7} (1-e^{-(7-i-b)})e^{-i}e^{-b}di\,db = \\
&= \int_{7-\ln(2)<i+b\le 7} (e^{-(i+b)}-e^{-7})di\,db = \\
&= \int_{7-\ln(2)<i+b\le 7} e^{-(i+b)}di\,db - \int_{7-\ln(2)<i+b\le 7} e^{-7}di\,db = \\
&= \int_{7-\ln(2)<i+b\le 7} e^{-(i+b)}di\,db - e^{-7}\left(\frac{7^2-(7-\ln(2))^2}{2}\right) \approx \\
&\approx \int_{i=0}^{7-\ln(2)} \int_{b=7-\ln(2)-i}^{7-i} e^{-(i+b)}db\,di - 0.0042 = \\
&= \int_{i=0}^{7-\ln(2)} \left(-e^{-(i+b)}|_{b=7-\ln(2)-i}^{7-i}\right) di - 0.0042 = \\
&= \int_{i=0}^{7-\ln(2)} e^{-7}di - 0.0042 = (7-\ln(2))e^{-7} - 0.0042 \approx 0.00155
\end{aligned}$$

Kokku klassifitseerija risk on seega $R(g^*) \approx 0.018 + 0.001 = 0.018$, s.t. vea tõenäosus on alla 2%.
Alati on mõtet võrrelda saadud klassifitseerija riski konstantse klassifitseerija

$g_c(I,B) \equiv 1$ riskiga:

$$
\begin{aligned}
R(g_c) &= \int_{I\times B} P(Y=2\,|\,I,B)dF(I,B) = \int P(E>7-I-B\,|\,I,B)dF(I,B) = \\
&= \int_{I+B\le 7} P(E>7-I-B\,|\,I,B)dF(I,B) + \\
&\qquad\qquad \int_{I+B>7} P(E>7-I-B)dF(I,B) = \\
&= \int_{I+B\le 7} e^{-7} di\,db + \int_{I+B>7} e^{-i-b} di\,db = \\
&= e^{-7}\frac{7^2}{2} + \left(1 - \int_{I+B\le 7} e^{-i-b} di\,db\right) \approx \\
&\approx 0.022 + 1 - \int_{i=0}^{7}\int_{b=0}^{7-i} e^{-i-b} db\,di = \\
&= 1.022 - \int_{i=0}^{7} \left(-e^{-b-i}|_{b=0}^{7-i}\right) di = \\
&= 1.022 - \int_{i=0}^{7} (e^{-i} - e^{-7}) di = 1.022 + 7e^{-7} - \left(-e^{-i}|_{i=0}^{7}\right) \approx \\
&\approx 1.029 - 1 + e^{-7} \approx 0.0296
\end{aligned}
$$

Niiet konstantse klassifitseerija risk ei ole oluliselt suurem.

Ülesanne 6: 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.

Lahendus: Eraldavaks jooneks on punktide $\mathbf{x}$ hulk, mis rahuldavad võrrandit:

$$
\frac{1}{2\pi\sqrt{\det(\mathbf{\Sigma})}} \exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu}_1)^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)}{2}\right) = \frac{1}{2\pi\sqrt{\det(\mathbf{\Sigma})}} \exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu}_2)^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2)}{2}\right)
$$

ehk

$$
(\mathbf{x}-\boldsymbol{\mu}_1)^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_1) = (\mathbf{x}-\boldsymbol{\mu}_2)^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2)
$$

avame mõlemal pool sulud ja taandame ruutliikme $\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x}$. Alles jääb esimese astme võrrand:

$$
\boldsymbol{\mu}_1^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_1 - 2\boldsymbol{\mu}_1^T\mathbf{\Sigma}^{-1}\mathbf{x} = \boldsymbol{\mu}_2^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_2 - 2\boldsymbol{\mu}_2^T\mathbf{\Sigma}^{-1}\mathbf{x}
$$

Seega eraldusjoon on sirgjoon. Huvi pärast viime võrrandit ilusamale kujule:

$$2(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{-1} \mathbf{x} = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_2 + \boldsymbol{\mu}_1)$$

$$(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{-1} \mathbf{x} = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{-1} \left(\frac{\boldsymbol{\mu}_2 + \boldsymbol{\mu}_1}{2} \right)$$

$$\mathbf{w}^T (\mathbf{x} - \boldsymbol{\mu}_c) = 0, \qquad \mathbf{w} = \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1), \qquad \boldsymbol{\mu}_c = \frac{\boldsymbol{\mu}_2 + \boldsymbol{\mu}_1}{2}$$

Seega eraldav joon läbib jaotuste keskväärtuste keskmist ning on ortogonaalne neid ühendava joonega, kui ortogonaalsuse all mõista omadust $\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{y} = 0$.

Ülesanne 7: 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?

Lahendus: Praktikas saab muidugi jaotust hinnata andmetest erinevatel viisidel. Võimalikud variandid:

- Oletada et klassid on konkreetsete jaotustega ning hinnata jaotuste parameetreid. Näiteks võime oletada klasside normaalsust ning hinnata nende jaotuste keskväärtusi ja kovariatsioonimaatrikse. Kui me oletame et klasside kovariatsioonimaatriksid on võrdsed, saame me klassifitseerijat mida nimetatakse *Fisheri diskriminandiks*

- Hinnata jaotust histogrammi abil, näiteks interpoleerides seda. Sellist klassifitseerijat nimetatakse tihti *Parzeni klassifitseerijaks.*

- Oletada tunnuste sõltumatust, hinnata iga tunnuse jaotust ning esitada ühisjaotust tunnuste jaotuste korrutisena. See on nn *Naïve Bayes-i klassifitseerija.* Kui me lubame teatud tunnuste vahel sõltuvusi piiratud kogusel, saame me *Bayesi võrgu.*