Neurovõrgud. Praktikum 8.

Ülesannete lahendused

Ülesanne 2: Näita et $\mathbf{w}$, mis maksimiseerib $J(\mathbf{w})$ avaldub kujul $\mathbf{w} = \lambda\hat{\mathbf{\Sigma}}^{-1}(\hat{\boldsymbol{\mu}}_2 - \hat{\boldsymbol{\mu}}_1)$ kus λ on suvaline reaalarv.

Lahendus: Tähistame $\mathbf{S} = (\hat{\boldsymbol{\mu}}_2 - \hat{\boldsymbol{\mu}}_1)(\hat{\boldsymbol{\mu}}_2 - \hat{\boldsymbol{\mu}}_1)^T$. Miinimumi otsimiseks paneme $\nabla J(\mathbf{w}) = 0$, millest saame

$$\begin{aligned} 2\mathbf{S}\mathbf{w}(\mathbf{w}^T\hat{\mathbf{\Sigma}}\mathbf{w}) - 2\hat{\mathbf{\Sigma}}\mathbf{w}(\mathbf{w}^T\mathbf{S}\mathbf{w}) &= 0 \\ \mathbf{S}\mathbf{w} &= \lambda\hat{\mathbf{\Sigma}}\mathbf{w}, \qquad \text{kus} \quad \lambda = \frac{\mathbf{w}^T\mathbf{S}\mathbf{w}}{\mathbf{w}^T\hat{\mathbf{\Sigma}}\mathbf{w}} \\ \hat{\mathbf{\Sigma}}^{-1}\mathbf{S}\mathbf{w} &= \lambda\mathbf{w} \end{aligned}$$

Kuna $\mathbf{S}\mathbf{w} = (\hat{\boldsymbol{\mu}}_2 - \hat{\boldsymbol{\mu}}_1)(\hat{\boldsymbol{\mu}}_2 - \hat{\boldsymbol{\mu}}_1)^T\mathbf{w} = k(\hat{\boldsymbol{\mu}}_2 - \hat{\boldsymbol{\mu}}_1)$ siis $\mathbf{w}$ peab olema vektori $\hat{\mathbf{\Sigma}}^{-1}(\hat{\boldsymbol{\mu}}_2 - \hat{\boldsymbol{\mu}}_1)$ kordne. MOTT.

Ülesanne 3: Fisheri diskriminant on ideaalne kui klasside kovariatsioonimaatriksid on võrdsed, mis praktikas ei pruugi kehtida. Kas seda saaks äkki mingi eeltöötluse abil parandada, näiteks normeerides ja tsentreerides muutujaid vms? Miks/Kuidas?

Lahendus: Kindlasti ei saa lineaarse teisendusega viia kovariatsioonimaatriksid võrdseteks. Mittelineaarse kujutisega oleks see võimalik, kuid kohe laest ilusa meetodi selle jaoks mina ise pakkuda ei oskaks. Igastahes oleks see omaette masinõppimise algoritm. Mõnikord võib selliseks teisenduseks teatud „loomulik" koordinaatide teisendus. Näiteks võib konstrueerida näiteid kus polaarkoordinaatides on klasside kovariatsioonide maatriksid võrdsed samas kui tavalistes koordinaatides mitte.

Ülesanne 4: Olgu kaks klassi jaotatud normaalselt sama kovariatsioonimaatriksiga (ning nad ei pea olema võrdtõenäosed, s.t. mitte tingimata $P(Y = 1) = P(Y = 2)$). Näita, et iga klassi *a-posteriori* tõenäosus avaldub logistilise funktsioonina, s.t.

$$P(Y = 1 \,|\, \mathbf{x}_0) = \frac{1}{1 + \exp(\mathbf{w}^T \mathbf{x}_0 + b)}$$

Lahendus:

$$\begin{aligned} P(Y = 1 \,|\, \mathbf{x}_0) &= \frac{P(\mathbf{x}_0 \,|\, Y = 1) P(Y = 1)}{P(\mathbf{x}_0)} = \\ &= \frac{P(\mathbf{x}_0 \,|\, Y = 1) P(Y = 1)}{P(\mathbf{x}_0 \,|\, Y = 1) P(Y = 1) + P(\mathbf{x}_0 \,|\, Y = 2) P(Y = 2)} = \\ &= \frac{1}{1 + \frac{P(\mathbf{x}_0 \,|\, Y=2) P(Y=2)}{P(\mathbf{x}_0 \,|\, Y=1) P(Y=1)}} \end{aligned}$$

kus

$$\frac{P(\mathbf{x}_0 \,|\, Y = 2) P(Y = 2)}{P(\mathbf{x}_0 \,|\, Y = 1) P(Y = 1)} = C \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_2)\right)}{\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_1)\right)}$$

saab viia kujule $\exp(\mathbf{w}^T \mathbf{x} + b)$.

Ülesanne $5\frac{1}{2}$ (1p): Nagu näha on, ei nõua k-nn klassifitseerija mingit „treenimist" — on vaja lihtsalt salvestada treeningandmeid. See eest on iga näite klassifitseerimine keerukas: tuleb vaadata kogu treeninghulga läbi ja otsida sealt naabreid. Kas saaks treenimisfaasis andmeid indekseerida selleks et pärast oleks võimalik leida lähimaid naabreid kiiremini? Miks/Kuidas?

Lahendus: Klassifitseerimisprotseduuri saaks võib-olla kiirendada kui näiteks eelnevalt jagada terve ruumi piirkondadeks, kus igas piirkonnas on vähemalt k treeningpunkti. Kui siis antud punkti järgi saab kiiresti leida teda sisaldavat piirkonda ja tema naabreid, siis piisab ainult nendes piirkondades punkte läbivaatamisest. Veel üks võimalus oleks andmete hierarhiline klasterdamine ning puu struktuurina hoidmine.

Ülesanne 6: Parzeni klassifitseerija on tegelikult Bayesi klassifitseerija empiiriline versioon. Seleta miks. Kas see klassifitseerija on sama mis RBF võrk?

Lahendus: Kirjutame reeglit ümber:

$$f(\mathbf{x}) = \mathrm{sign}\left(\sum_{i=1}^{n} y_i K(\mathbf{x}, \mathbf{x}_i)\right) = \mathrm{sign}\left(\sum_{i, y_i = +1} y_i K(\mathbf{x}, \mathbf{x}_i) + \sum_{i, y_i = -1} y_i K(\mathbf{x}, \mathbf{x}_i)\right) =$$

$$= \mathrm{sign}\left(\sum_{i, y_i = +1} K(\mathbf{x}, \mathbf{x}_i) - \sum_{i, y_i = -1} K(\mathbf{x}, \mathbf{x}_i)\right) =$$

$$= \begin{cases} 1, \text{ kui } \hat{P}(1 \,|\, \mathbf{x}) > \hat{P}(-1 \,|\, \mathbf{x}) \\ -1, \text{ kui } \hat{P}(-1 \,|\, \mathbf{x}) > \hat{P}(1 \,|\, \mathbf{x}) \end{cases}$$

kus

$$\hat{P}(y \,|\, \mathbf{x}) = \sum_{i, y_i = y} K(\mathbf{x}, \mathbf{x}_i)$$

on klassi posterioorse tõenäosuse empiiriline hinnang.
Ei, see pole RBF võrk kuna siin ei valita koeffitsiente nagu seda tehakse RBF-i puhul.