Навчання з учителем. Комп'ютерне завдання

Святослав Хмелевський, ФІ-82

21 жовтня 2021 р.

Завдання: Програма на вхід приймає три скінченні послідовності двовимірних дійснозначних векторів. Перша послідовність містить значення, що з більшою ймовірністю належить певному (невідомому програмі) ґауссовому розподілу, ніж значення з другої послідовності. Ці дві послідовності програма має використати для навчання. Третя послідовність містить довільні вектори, і їх треба класифікувати. На вихід програма виводить індекси тих елементів третьої послідовності, що належать першому класу.

Розв'язок: Нехай $\tau = \{N(\boldsymbol{\mu}, \Sigma) : \ \boldsymbol{\mu} \in \mathbb{R}^2, \ \Sigma \in \mathbb{R}^{2\times 2}\}$ – клас двовимірних ґауссових розподілів, $\theta \in \mathbb{R}_+$ – порогове значення, $T \in (\mathbb{R}^2 \times \{-1, +1\})^m$ – наявний емпіричний матеріал і $(\boldsymbol{x}, k) \in T$.

У випадку класу двовимірних ґауссових розподілів, ймовірності $p(\boldsymbol{x})$ дорівнюватимуть $\dfrac{1}{2\pi\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$. Що буде використано.

$$p(\boldsymbol{x}) \lesseqgtr \theta \overset{p\in\tau}{\Longrightarrow} \frac{1}{2\pi\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})} \lesseqgtr \theta$$

Тепер прологарифмуємо і згрупуємо елементи отриманої нерівності.

$$-\frac{1}{2}\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x} + \boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{x} - \frac{1}{2}(\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu} + \ln(4\pi^2\theta^2\det(\Sigma))) \lesseqgtr 0$$

Розпишемо більш детально деякі моменти отриманої нерівність.

$$-\frac{1}{2}\sum_{i,j=1}^{2} \Sigma_{ij}^{-1}\boldsymbol{x}_i\boldsymbol{x}_j + \sum_{i,j=1}^{2} \Sigma_{ij}^{-1}\boldsymbol{\mu}_i\boldsymbol{x}_j - \frac{1}{2}(\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu} + \ln(4\pi^2\theta^2\det(\Sigma))) \lesseqgtr 0$$

Отриману нерівність можна записати в більш компактному вигляді, враховуючи особливості коваріаційної матриці Σ.

$$p(\boldsymbol{x}) \lesseqgtr \theta \implies \ldots \implies \boldsymbol{\alpha}^T \cdot \boldsymbol{y} \lesseqgtr 0,$$

де $\boldsymbol{\alpha} = \Bigg(-\frac{1}{2}(\boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu} + \ln(4\pi^2\theta^2 \det(\Sigma))), \sum_{i=1}^{2} \Sigma_{1i}^{-1} \boldsymbol{\mu}_i, \sum_{i=1}^{2} \Sigma_{2i}^{-1} \boldsymbol{\mu}_i,$
$-\frac{1}{2}\Sigma_{11}^{-1}, -\Sigma_{12}^{-1}, -\frac{1}{2}\Sigma_{22}^{-1} \Bigg)^T, \ \boldsymbol{y} = (1, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_1^2, \boldsymbol{x}_1\boldsymbol{x}_2, \boldsymbol{x}_2^2)^T.$
Тепер треба подумати про забезпечення виконання властивостей коваріаційної матриці Σ. Симетричність матриці уже врахована, залишається вирішити проблему з додатньою визначеністю матриці Σ. Для цього достатньо, щоб виконувалася додатня визначеність для матриці Σ^{-1}. Далі використаємо одну з властивість власних чисел.

$$\boldsymbol{v}^T \Sigma^{-1} \boldsymbol{v} > 0, \ \Sigma^{-1}\boldsymbol{v} = \lambda \boldsymbol{v} \implies \lambda \|\boldsymbol{v}\|^2 > 0 \implies \lambda > 0,$$

де $\boldsymbol{v}$ – власний вектор матриці Σ^{-1}, а λ – відповідне власне значення. Тобто $\boldsymbol{v} \in \{\boldsymbol{u} : \ \boldsymbol{u} \in \mathbb{R}^2, \ \Sigma^{-1}\boldsymbol{u} = \lambda \boldsymbol{u}, \ \lambda \in \mathbb{R}\}$. У випадку симетричної матриці, власні числа приймають дійсні значення. При цьому треба, щоб виконувалося $\boldsymbol{v}^T \Sigma^{-1} \boldsymbol{v} > 0$ для власних векторів матриці Σ^{-1}. Тобто, щоб $\boldsymbol{\alpha}^T \cdot (0, 0, 0, -\boldsymbol{v}_1^2, -\boldsymbol{v}_1\boldsymbol{v}_2, -\boldsymbol{v}_2^2)^T = \frac{1}{2}\boldsymbol{v}^T \Sigma^{-1} \boldsymbol{v} > 0$.
Таким чином, щоб забезпечити додатню визначеність матриці Σ^{-1} потрібно, щоб для векторів $(0, 0, 0, -\boldsymbol{v}_1^2, -\boldsymbol{v}_1\boldsymbol{v}_2, -\boldsymbol{v}_2^2)^T$ утворених з власних векторів виконувалися наведені вище співвідношення.

Візьмемо за основу алгоритм навчання персептрону.
Весь алгоритм буде мати наступний вигляд.

1. Ініціалізація:
   - 1.1 На вхід подається емпіричний матеріал $T \in (\mathbb{R}^2 \times \{-1, +1\})^m$ (m – кількість елементів).
   - 1.2 $\widetilde{T} = \{(\boldsymbol{y}, k) : \ \boldsymbol{y} = (1, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_1^2, \boldsymbol{x}_1\boldsymbol{x}_2, \boldsymbol{x}_2^2)^T, \ (\boldsymbol{x}, k) \in T\}$.
   - 1.3 $\boldsymbol{\alpha}^0 = \mathbf{0}, \ t = 0$.
2. Основний цикл:
   - 2.1 Якщо $\exists (\boldsymbol{y}, k) \in \widetilde{T} : \ (\boldsymbol{\alpha}^t)^T \cdot k\boldsymbol{y} \le 0$. Тоді:
     - 2.1.1 $\boldsymbol{\alpha}^{t+1} = \boldsymbol{\alpha}^t + k\boldsymbol{y}$.
     - 2.1.2 $t \leftarrow t + 1$, перехід до 2.
   - 2.2 Якщо $\forall (\boldsymbol{y}, k) \in \widetilde{T} : \ (\boldsymbol{\alpha}^t)^T \cdot k\boldsymbol{y} > 0$. Тоді:
     - 2.2.1 $\widetilde{T} \leftarrow \widetilde{T} \cup \{((0, 0, 0, -\boldsymbol{v}_1^2, -\boldsymbol{v}_1\boldsymbol{v}_2, -\boldsymbol{v}_2^2)^T, 1) : \Sigma^{-1}\boldsymbol{v} = \lambda\boldsymbol{v}, \ \boldsymbol{v} \in \mathbb{R}^2, \ \lambda \in \mathbb{R}\}$.
     - 2.2.2 Якщо $\exists (\boldsymbol{y}, k) \in \widetilde{T} : \ (\boldsymbol{\alpha}^t)^T \cdot k\boldsymbol{y} \le 0$. Тоді: Перехід до кроку 2.1.1.
     - 2.2.3 Якщо $\forall (\boldsymbol{y}, k) \in \widetilde{T} : \ (\boldsymbol{\alpha}^t)^T \cdot k\boldsymbol{y} > 0$. Тоді: Робота алгоритму завершена.