Modyfikacje LDA

Na podstawie rozdziałów 4.3.1, 4.3.2 i 4.3.3 książki *The Elements of Statistical Learning* (Trevor Hastie, Robert Tibshirani, Jerome Friedman)

Mateusz Łyczko

Instytut Matematyczny UWr

21.04.2023

Regularized Discriminant Analysis - RDA (1)

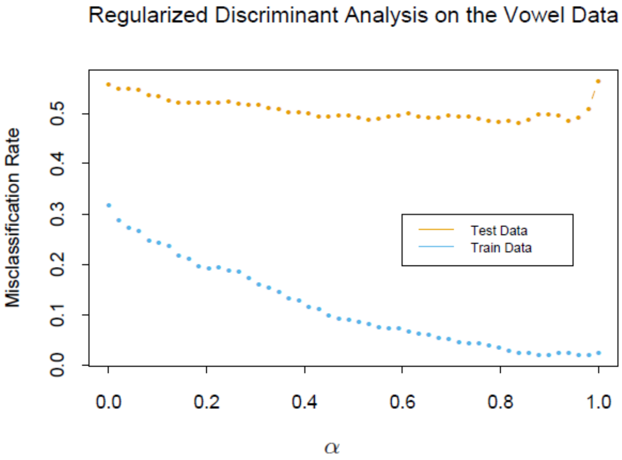
To kompomis między LDA (Linear Discriminant Analysis) a QDA (Quadratic Discriminant Analysis). Macierze mają tutaj taką postać:

$$\hat{\mathbf{\Sigma}}_k(\alpha) = \alpha\hat{\mathbf{\Sigma}}_k + (1-\alpha)\hat{\mathbf{\Sigma}}.$$

Wartość α należy wybrać, może być ona z przedziału $[0, 1]$.

Regularized Discriminant Analysis - RDA (2)

Poniżej przedstawiamy rysunek uzyskany dla pewnych danych, gdy zmieniamy α:



Regularized Discriminant Analysis - RDA (3)

RDA można jeszcze zmodyfikować, zastępując $\widehat{\Sigma}$ przez

$$\hat{\mathbf{\Sigma}}(\gamma) = \gamma\hat{\mathbf{\Sigma}} + (1-\gamma)\hat{\sigma}^2\mathbf{I}$$

Wartość γ również jest z przedziału [0,1].

Obliczenia dla LDA i QDA

Okazuje się, że obliczenia dla tych metod dają się uprościć, gdy zastosujemy diagonalizację dla macierzy w nich występujących. Skupimy się tutaj na QDA. Z nią związany jest taki wzór:

$$\delta_k(x) = -\frac{1}{2}\log|\mathbf{\Sigma}_k| - \frac{1}{2}(x-\mu_k)^T\mathbf{\Sigma}_k^{-1}(x-\mu_k) + \log\pi_k.$$

Używając diagonalizacji można zamienić pewne obiekty w następujący sposób:

- $(x-\hat{\mu}_k)^T\hat{\mathbf{\Sigma}}_k^{-1}(x-\hat{\mu}_k) = [\mathbf{U}_k^T(x-\hat{\mu}_k)]^T\mathbf{D}_k^{-1}[\mathbf{U}_k^T(x-\hat{\mu}_k)]$;
- $\log|\hat{\mathbf{\Sigma}}_k| = \sum_\ell \log d_{k\ell}$.

Reduced Rank Linear Discriminant Analysis (1)

To LDA, które jest zredukowane do przestrzeni niższego wymiaru. Szuka się tutaj zmiennych dyskryminacyjnych (discriminant variables lub canonical variables) i wybiera później pewną ich liczbę. Aby je znaleźć, postępujemy wg takiego schematu:
1. Szukamy macierzy M przechowującej centroidy klas (jest rozmiaru $K \times p$, K to liczba klas, p to wymiar danych) oraz wspólnej dla klas macierzy kowariancji W.
2. Liczymy $M^* = MW^{-1/2}$ korzystając z diagonalizacji W.
3. Znajdujemy macierz B^*, która jest macierzą kowariancji M^* i diagonalizujemy ją: $B^* = V^* D_B (V^*)^T$.
Wtedy m-ta zmienna dyskryminacyjna to $Z_m = v_m^T X$, gdzie $v_m = W^{-1/2} v_m^*$, a gdzie z kolei v_m^* to m-ta kolumna V^*.

Reduced Rank Linear Discriminant Analysis (2)

Fisher podszedł do tego problemu inaczej. Chcemy zmaksymalizować tzw. iloraz Rayleigha (Rayleigh quotient):

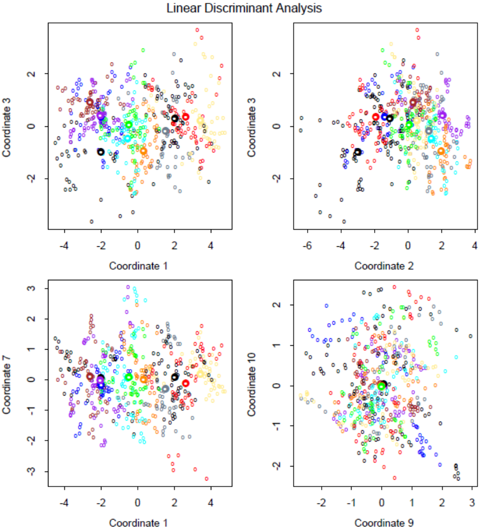
$$\max_a \frac{a^T \mathbf{B} a}{a^T \mathbf{W} a}$$

Równoważnie:

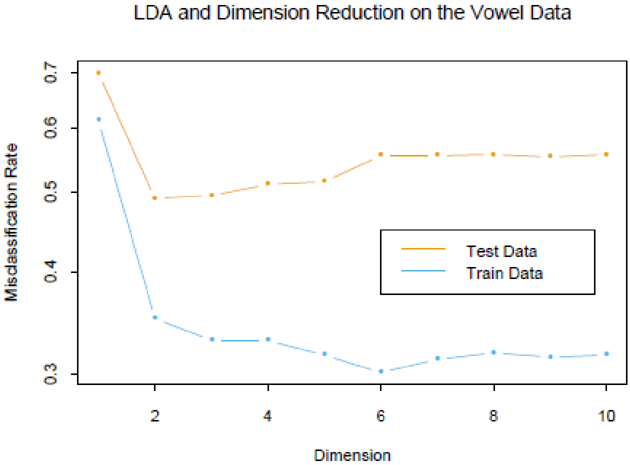
$$\max_a a^T \mathbf{B} a \text{ subject to } a^T \mathbf{W} a = 1.$$

Okazuje się, że wyniki uzyskane tą metodą są dokładnie takie same jak te wyliczone poprzednim sposobem.

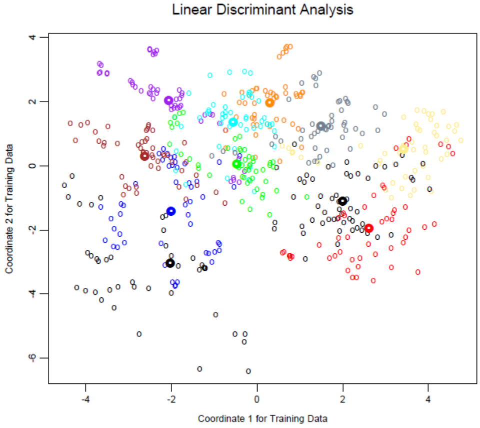
Reduced Rank Linear Discriminant Analysis, Rysunek 1



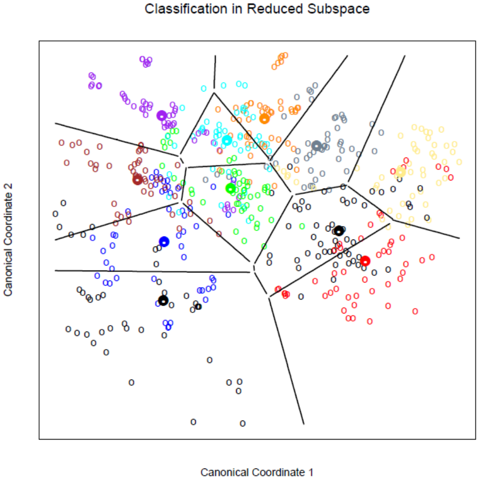
Reduced Rank Linear Discriminant Analysis, Rysunek 2



Reduced Rank Linear Discriminant Analysis, Rysunek 3



Reduced Rank Linear Discriminant Analysis, Rysunek 4



Bibliografia

1. Trevor Hastie, Robert Tibshirani, Jerome Friedman, *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*