Projekt – klasyfikator LDA

Celem projektu jest stworzenie klasyfikatora dyskryminacji liniowej LDA, oraz pokazanie jego skuteczności na różnych zbiorach danych.

Sposób działania LDA opiera się na tym że posiadamy pewną ilość klas, w tym przypadku dwie, i otrzymawszy pewną liczbę cech danego obiektu klasyfikator powinien przypisać go do danej klasy.

Idea klasyfikacji opiera się na założeniu normalności rozkładu danej klasy, to znaczy, obiekty danej klasy mają rozkład normalny. Dla wyższych niż 1 wymiar gęstość prawdopodobieństwa wyraża się wzorem:

*p*(**x**;*μ*,Σ)=(2*π*)^(n/2)\*​∣Σ|^(1/2)\*​​exp(−21​(**x**−*μ*)*T*Σ−1(**x**−*μ*))

Gdzie Σ to macierz covariancji a *μ* to średnia dla danej klasy.

Ważnym założeniem LDA jest fakt że obydwie klasy mają jednakową macierz covariancji, co sprawia że granica decyzyjna jest hiperpłaszczyzną.

Porównując gęstości obydwóch klas otrzymujemy granicę decyzyjną:

X\*Σ^(-1)\*( *μ1- μ2)>=* Σ^(-1)\*( *μ1- μ2)( μ1+ μ2)/2* {\displaystyle {\vec {w}}=\Sigma ^{-1}({\vec {\mu }}\_{1}-{\vec {\mu }}\_{0})}{\displaystyle c={\vec {w}}\cdot {\frac {1}{2}}({\vec {\mu }}\_{1}+{\vec {\mu }}\_{0})}

2.

Dane wykorzystane w eksperymencie pochodzą z 2 źródeł, jedne są sztucznie wygenerowanymi danymi spełniającymi warunki LDA, drugi pochodzą z targu rybnego. Zadanie klasyfikatora ma być podanie gatunku ryby gdy znane są jej wymiary. Istniją silne przesłanki o normalności takiego rozkładu gdyż ryby w ramach jednego gatunku powinny mieć rozmiary zbliżone do średniej i odychlenia od niej powinny być dobrze modelowane przez rozkład normalny.

Równe macierze covariancji również sa prawdopodobne gdyż rozmiary ryb powinny zachowywać pewne podobieństwo w odchylaniu od średniej z powodu że są to podobne zwierzęta.

Wyniki eksperymentów przedstawione są za pomocą odpowednich wskaźników w pliku z impementacją (F1, recal, precision, accuracy)

4.

Klasyfikator działa bardzo dobrze przy danych z rozkładu normalnego jeżeli ich średnie wartości są od siebie oddalone oraz wartości variancji nie są zbyt duże w przeciwnym wypadku istnieje duże prawdopodobieństwo że próbki się ze sobą wymieszają. Klasyfkiator sprawdza się również bardzo dobrze w klasyfikacji ryb do gatunków z powodu podanego powyżej, wymiary ryb mają rozkład zbliżony do normalnego.