Autoři: RNDr. Eva Janoušová, Mgr. et Mgr. Petr Dluhoš
Institut biostatistiky a analýz, Masarykova univerzita, Brno

Fisherova lineární diskriminace – příklad

Bylo provedeno měření objemu hipokampu a amygdaly (v cm^3) u 3 pacientů s Alzheimerovou chorobou ( ) a 3 kontrolních subjektů ( ) (označení D – diseased, H – healthy). Naměřené hodnoty objemu hipokampu a amygdaly u pacientů (resp.) a kontrol (resp.) byly zaznamenány do matic resp. :

Určete, zda testovací subjekt patří do skupiny pacientů či kontrolních subjektů pomocí Fisherovy lineární diskriminace.

Řešení:

Principem Fisherovy lineární diskriminace je transformace do jednorozměrného (1D) prostoru tak, že chceme maximalizovat vzdálenost skupin (odráží se v čitateli Fisherova diskriminačního kritéria) a minimalizovat variabilitu uvnitř skupin (odráží se ve jmenovateli Fisherova diskriminačního kritéria). Fisherovo diskriminační kritérium je tedy ve tvaru:

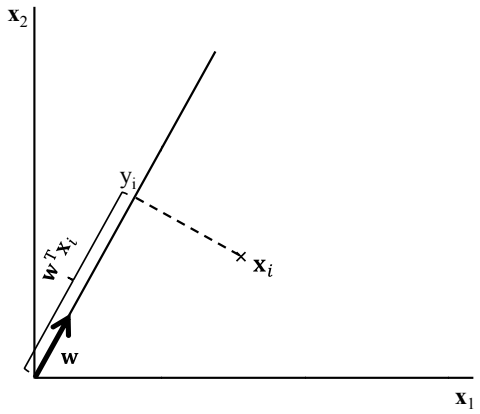
_____

kde je projekce centroidu pacientů do 1-D prostoru, je projekce centroidu kontrol , je rozptyl uvnitř třídy pacientů po projekci do 1-D prostoru a je rozptyl uvnitř třídy kontrol. Centroidy jsou vícerozměrné průměry pro třídu pacientů a kontrol:

— — — ,

— — — ,

kde je hodnota první proměnné u -tého subjektu a je počet proměnných. Projekce centroidů do 1-D prostoru mohou být vypočítány jako a , kde je váhový vektor udávající směr 1-D prostoru, do něhož promítáme. Obecně může být průmět jakéhokoliv bodu do 1D prostoru vypočítán jako a znázorněn pomocí *Obrázku 1*.



Obrázek 1. Znázornění projekce bodu do 1-D prostoru daného směrovým vektorem **w**. Bod reprezentuje -tý subjekt a je jeho projekce. Osy a odpovídají dvěma proměnným.

Rozptyl uvnitř třídy pacientů po projekci do 1-D prostoru ( ) lze vypočítat jako čtverec vzdáleností projekcí bodů odpovídajících jednotlivým pacientům od projekce centroidu:

kde je kovarianční matice pacientů. Obdobně je možné rozptyl uvnitř třídy kontrol po projekci do 1-D prostoru ( ) vypočítat jako:

kde je kovarianční matice kontrol.

Dále si rozepíšeme součet rozptylů uvnitř jednotlivých tříd po transformaci do 1D prostoru, který se vyskytuje ve jmenovateli Fisherova diskriminačního kritéria:

kde je suma čtverců variability uvnitř skupin a lze ji vypočítat jako: . V obecném případě, kdy nejsou vyvážené počty subjektů ve skupinách, se počítá vážená suma čtverců variability uvnitř skupin jako .

Čitatel Fisherova diskriminačního kritéria si můžeme rozepsat jako:

kde je suma čtverců variability mezi skupinami.

Fisherovo diskriminační kritérium tedy můžeme vyjádřit jako:

Chceme maximalizovat , proto zderivujeme a položíme výraz roven 0:

Víme, že má směr , protože , kde je nějaký skalár. U vektoru nás nezajímá jeho modul (tzn. velikost), jen jeho směr, proto můžeme pominout skalární členy a . Dostáváme tedy:

Po odvození vzorečku pro výpočet váhového vektoru do něj můžeme dosadit konkrétní hodnoty centroidů (vícerozměrných průměrů) pro třídu pacientů a kontrol, tzn. , .
Pro výpočet sumy čtverců variability mezi skupinami využijeme výběrové kovarianční matice a (výpočet vícerozměrných průměrů a výběrových kovariančních matic lze nalézt ve Cvičení 1). Suma čtverců variability mezi skupinami bude tedy spočítána jako a její inverze jako .

Váhový vektor (diskriminační směr) poté tedy můžeme spočítat následujícím způsobem:

Protože nás nezajímá modul váhového vektoru, ale jen jeho směr, můžeme váhový vektor přeškálovat na: . Nyní můžeme vypočítat průměty centroidů do 1D prostoru:

A následně vypočteme průmět hraničního bodu v 1D prostoru:

Hraniční bod lze vypočítat i takto: − −

− − − − ——

— (protože jsme váhový vektor přeškálovali pomocí vynásobení −, musíme vynásobit i a pak získáváme -31).

Pokud chceme zařadit nový subjekt do jedné z daných tříd, musíme nejprve vypočítat jeho průmět do 1-D prostoru:

Průmět následně srovnáme s hraničním bodem: protože , subjekt zařadíme do skupiny kontrolních subjektů (kontrolní subjekty leží nalevo od hraničního bodu, protože centroid kontrolních subjektů má menší (=více negativní) hodnotu než hraniční bod).

Po výpočtu váhového vektoru a hraničního bodu můžeme určit obecnou rovnici hranice (normálou hraniční přímky je váhový vektor):

Pro vykreslení hranice je vhodné vyjádřit hranici ve tvaru:
Nová osa, do níž se promítá, má směr odpovídající váhovému vektoru (je kolmá k hranici) a prochází počátkem a lze ji tedy vyjádřit obecnou rovnicí jako:

Pokud nás zajímají souřadnice hraničního bodu v původním prostoru, využijeme znalosti, že hraniční bod je průsečík hranice a nové osy:

—

Souřadnici pak vypočítáme z druhé rovnice jako: — ——

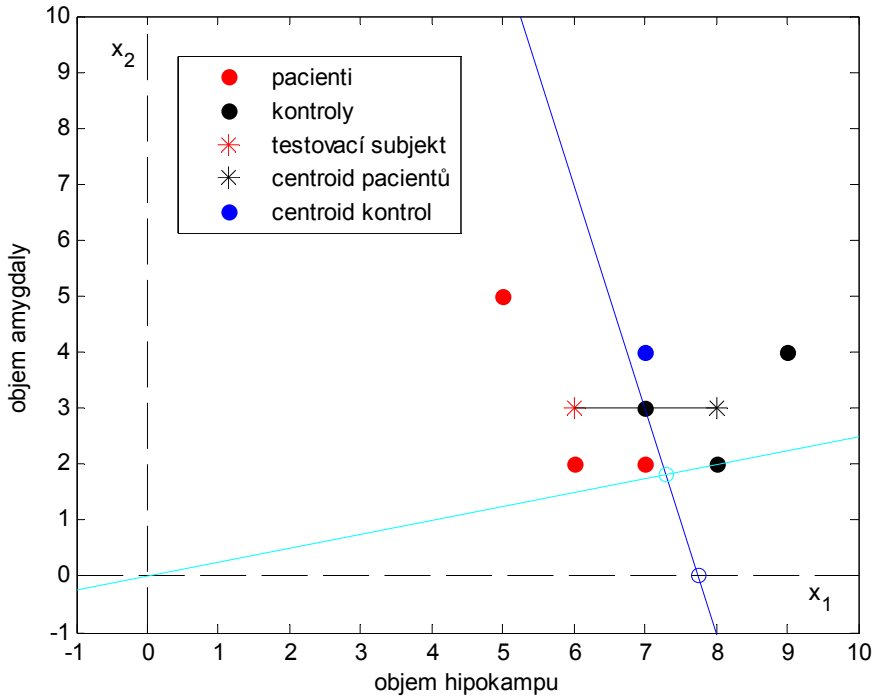
Souřadnice hraničního bodu v původním prostoru jsou tedy: —— —

Ověření, že po projekci hraničního bodu dostanu hodnotu -31: —

—

—— — ——

Klasifikaci pomocí Fisherovy lineární diskriminační analýzy si na závěr znázorníme pomocí *Obrázku 2*.



Obrázek 2. Znázornění klasifikace pomocí Fisherovy lineární diskriminační analýzy. Klasifikační hranice je znázorněna tmavě modře, nová osa, do níž se promítá, světle modře a hraniční bod je vyznačen tmavě modrým prázdným kolečkem. Původní osy a odpovídající dvěma proměnným (objemu hipokampu a amygdaly) jsou znázorněny čárkovanými čarami.

Poznámka: Pokud bychom váhový vektor znormovali, hraniční bod by přímo ležel ve vzdálenosti od počátku:

(tzn. hraniční bod leží ve vzdálenosti od počátku v původních souřadnicích)