Zestaw 2 - Zadanie 7

Metody probabilistyczne w uczeniu maszynowym

Łukasz Trzos

Treść zadania

Wykaż, że estymator najmniejszych kwadratów dla modelu regresji liniowej z szumem gaussowskim $\mathcal{N}(0, \sigma^2)$ jest najlepszym (czyli mającym najmniejszą kowariancję) estymatorem spośród wszystkich liniowych nieobciążonych estymatorów.

Przyjmujemy, że macierz A jest nie większa od macierzy B (piszemy $A \leqslant B$), gdy $B - A = C$ dla pewnej półdodatnio określonej macierzy C.

Rozwiązanie

Estymator najmniejszych kwadratów obliczamy ze wzoru:

$$\hat{\theta}^{ls} = (X^T X)^{-1} X^T y$$

gdzie X oraz y są zdefiniowane w zestawie zadań.
Chcemy pokazać, że jest on lepszy od dowolnego innego estymatora nieobciążonego. Niech $\bar{\theta}$ będzie innym estymatorem, możemy go zapisać w postaci $\bar{\theta} = Cy$, gdzie $C = (X^T X)^{-1} X^T + D$.

Z zadania nr. 5 mamy:

$$Cov(\bar{\theta}) = Cov(Cy) = CCov(y)C^T = CCov(X\theta + \epsilon)C^T$$

Mamy $y = X\theta + \epsilon$, gdzie ϵ jest wektorem niezależnych zmiennych o rozkładzie normalnym. $X\theta$ jest stałe, zatem:

$$CCov(X\theta + \epsilon)C^T = CCov(\epsilon)C^T = C\sigma^2 IC^T = \sigma^2 CC^T$$

$Cov(\epsilon) = \sigma^2 I$, ponieważ kowariancja zmiennej o rozkładzie $\mathcal{N}(0, \sigma^2)$ z samą sobą wynosi σ^2, a kowariancja 2 niezależnych zmiennych jest równa 0. Rozwijamy C zgodnie z definicją:

$$\sigma^2 CC^T = \sigma^2((X^T X)^{-1} X^T + D)((X^T X)^{-1} X^T + D)^T = \sigma^2((X^T X)^{-1} X^T + D)((((X^T X)^{-1} X^T)^T + D^T) =$$

$$= \sigma^2((X^T X)^{-1} X^T + D)(X((X^T X)^{-1} + D^T) = \sigma^2((X^T X)^{-1} + (X^T X)^{-1}(DX)^T + (DX)(X^T X)^{-1} + DD^T)$$

DX = 0

Chcemy teraz pokazać, że estymator jest nieobciążony tylko wtedy, gdy $DX = 0$. W tym celu obliczamy jego wartość oczekiwaną:

$$E[\bar{\theta}] = E[Cy] = CE[y] = ((X^T X)^{-1} X^T + D)E[X\theta + \epsilon] = ((X^T X)^{-1} X^T + D)E[X\theta] + ((X^T X)^{-1} X^T + D)E[\epsilon] =$$

$$= ((X^T X)^{-1} X^T + D)(X\theta) = ((X^T X)^{-1} X^T X + DX)\theta = (I + DX)\theta$$

Po drodze korzystamy z faktów, że $X\theta$ jest stałe, a $E[\epsilon] = 0$, bo jest to wektor zmiennych o rozkładzie $\mathcal{N}(0, \sigma^2)$

Z definicji estymatora nieobciążonego: $E[\bar{\theta}] = \theta$, zatem

$$(I + DX)\theta = \theta + DX\theta = \theta$$

co dla niezerowego θ oznacza, że $DX = 0$

Kontynuujemy obliczenia kowariancji estymatora:

$$Cov(\bar{\theta}) = \sigma^2((X^TX)^{-1} + (X^TX)^{-1}(DX)^T + (DX)(X^TX)^{-1} + DD^T = \sigma^2((X^TX)^{-1} + DD^T) =$$

$$\sigma^2(X^TX)^{-1} + \sigma^2 DD^T = Cov(\hat{\theta}^{ls}) + \sigma^2 DD^T$$

Korzystamy z zad.6 w którym policzyliśmy kowariancję estymatora najmniejszych kwadratów w modelu. Wiemy, że macierz postaci DD^T jest zawsze dodatnio półokreślona. Zatem doszliśmy do równości:

$$Cov(\bar{\theta}) - Cov(\hat{\theta}^{ls}) = C'$$

dla pewnej dodatnio półokreślonej macierzy C'. Estymator najmniejszych kwadratów ma więc nie większą kowariancję od dowolnego innego estymatora nieobciążonego. $\square$