Metody liniowe regresji

Bartosz Zimończyk

Marzec 2021

Seminarium magisterskie dla specjalizacji analiza danych

O czym będziemy dzisiaj rozmawiać

Regresja liniowa

Mając wektor zmiennych objaśniających $X = (X_1, \dots, X_p)^T$ chcemy mieć metodę przewidywania zmiennej objaśnianej Y. Model regresji liniowej ma formę

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j,$$

gdzie współczynniki β_j są nieznane. Na później chcemy mieć wygodniejszy zapis. Niech $\mathbf{X}$ będzie macierzą $N \times (p+1)$, gdzie każdy wiersz to wektor zmiennych objaśniających, a $\mathbf{y}$ to wektor ze zmiennymi objaśnianymi, a β to wektor współczynników β_j. Zakładamy, że

$$Y = f(X), \quad \varepsilon \sim N(0, \sigma^2).$$

Regresja liniowa

Najpopularniejszeą metodą wyznaczania β jest metoda najmniejszych kwadratów, tj. chcemy znaleźć takie β_j, które minimalizują poniższą wartość

$$\begin{aligned} \mathrm{RSS}(\beta) &= \sum_{i=1}^{N} (y_i - f(x_i))^2 \\ &= \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=i}^{p} x_{ij}\beta_j)^2. \end{aligned}$$

Rozwiązaniem jest wektor

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}.$$

Regresja liniowa

- Estymator σ^2 to $\hat{\sigma}^2 = \frac{1}{N-p-1}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$
- $\hat{\beta} \sim N(\beta, \mathbf{X}^T\mathbf{X})^{-1}\sigma^2)$
- $(N-p-1)\hat{\sigma}^2 \sim \sigma^2\chi^2_{N-p-1}$

Testowanie istotności współczynników

Aby przetestować hipotezę $H_{0j} : \beta_j = 0$ przeciwko $H_{1j} : \beta_j \neq 0$, możemy użyć statystyki

$$Z_j = \frac{\hat{\beta}_j}{\hat{\sigma}\sqrt{v_j}},$$

gdzie v_j to $j$ty element przekątnej macierzy $(\mathbf{X}^T\mathbf{X})^{-1}$. Przy hipotezie Z_j ma rozkład t_{N-p-1}.

Testowanie istotności współczynników

Czasami chcemy sprawdzić czy zwiększenie liczby współczynników statystycznie istotnie coś zmienia. Do tego możemy wykorzystać statystykę F

$$F = \frac{(\mathrm{RSS}_0 - text RSS_1)/(p_1 - p_0)}{text RSS_1/(N - p - 1)},$$

gdzie RSS_1 jest rezydualną sumą kwadratów większego modelu z $p_1 + 1$ współczynnikami, a RSS_0 mniejszego, z $p_0 + 1$ współczynnikami. Przy hipotezie, że nie ma różnicy pomiędzy modelami, statystyka F ma rozkład $F_{p_1 - p_0, N - p_1 - 1}$.

Przykład

TABLE 3.2. *Linear model fit to the prostate cancer data. The Z score is the coefficient divided by its standard error (3.12). Roughly a Z score larger than two in absolute value is significantly nonzero at the $p = 0.05$ level.*

Term	Coefficient	Std. Error	Z Score
`Intercept`	2.46	0.09	27.60
`lcavol`	0.68	0.13	5.37
`lweight`	0.26	0.10	2.75
`age`	−0.14	0.10	−1.40
`lbph`	0.21	0.10	2.06
`svi`	0.31	0.12	2.47
`lcp`	−0.29	0.15	−1.87
`gleason`	−0.02	0.15	−0.15
`pgg45`	0.27	0.15	1.74

Rysunek 1: Model regresji liniowej dopasowany do danych raka prostaty.

Przykład

We wcześniejszej tabeli możemy zauważyć, że mamy cztery statystycznie nieistotne współczynniki. Zatem policzmy statystykę F

$$F = \frac{(32.81 - 29.43)/(9-5)}{29.43/(67-9)} = 1.67,$$
$$\mathrm{P}(F_{4,58} > 1.67) = 0.17.$$

Zatem nie mamy przypuszczeń, aby odrzucić hipotezę zerową mówiącą, że są różnice pomiędzy tymi modelami.

Wybór najlepszego modelu

Tworząc model regresji, chcemy znaleźć taki podzbiór wszystkich współczynników który jest najbardziej optymalny. Możemy to porównywać za pomocą różnych metryk:

- RSS,
- AIC,
- BIC,
- RIC

i wiele innych, które są modyfikacjami powyższych. Przyjmujemy, że chcemy minimalizować te metryki.

Algorytmiczne metody wyboru modelu

- *Forward-stepwise selection* - zaczynamy budować model od samego interceptu i w każdym kroku dodajemy zmienną która minimalizuje wcześniej wybraną metrykę dopasowania modelu.
- *Backward-stepwise selection* - zaczynamy od pełnego modelu i wyrzucamy zmienną która daje największą wartość wcześniej wybranej metryki.

Przykład

TABLE 3.3. *Estimated coefficients and test error results, for different subset and shrinkage methods applied to the prostate data. The blank entries correspond to variables omitted.*

Term	LS	Best Subset	Ridge	Lasso	PCR	PLS
Intercept	2.465	2.477	2.452	2.468	2.497	2.452
lcavol	0.680	0.740	0.420	0.533	0.543	0.419
lweight	0.263	0.316	0.238	0.169	0.289	0.344
age	−0.141		−0.046		−0.152	−0.026
lbph	0.210		0.162	0.002	0.214	0.220
svi	0.305		0.227	0.094	0.315	0.243
lcp	−0.288		0.000		−0.051	0.079
gleason	−0.021		0.040		0.232	0.011
pgg45	0.267		0.133		−0.056	0.084
Test Error	0.521	0.492	0.492	0.479	0.449	0.528
Std Error	0.179	0.143	0.165	0.164	0.105	0.152

Rysunek 2: Model różncyh typów regresji dopasowany do danych raka prostaty.

Regresja grzbietowa

Przeprowadzając regresję grzbietową chcemy znaleźć takie współczynniki β które minimalizują poniższe wyrażenie

$$\hat{\beta}^{\text{ridge}} = \arg\min_{\beta} \left\{ \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}.$$

Wyrażenie $\lambda \sum_{j=1}^{p} \beta_j^2$ odpowiada za *ściąganie* współczynników do 0. Parametr λ to *siła* z jaką chcemy zmniejszać β. W przypadku zapisu macierzowego estymator β ma rozwiązanie postaci

$$\hat{\beta}^{\text{ridge}} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}.$$

Regresja LASSO

Przeprowadzając LASSO chcemy znaleźć takie współczynniki β które minimalizują poniższe wyrażenie

$$\hat{\beta}^{\text{ridge}} = \arg\min_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}.$$

Wyrażenie $\lambda \sum_{j=1}^{p} |\beta_j|$ odpowiada za *ściąganie* współczynników do 0. Parametr λ to *siła* z jaką chcemy zmniejszać β. Przez to, że mamy wartość bezwzględną, ciężko jest wyliczyć analitycznie rozwiązanie. Jeśli jednak założymy, że $\mathbf{X}^T\mathbf{X} = \mathbf{I}$, to estymator β w regresji LASSO ma postać

$$\hat{\beta}^{\text{LASSO}} = (1 + N\lambda)^{-1}\hat{\beta}^{\text{LS}}.$$

Regresja ElasticNet

O *Elastic Net* możemy myśleć jak o połączeniu regresji grzbietowej i LASSO. W tym przypadku minimalizujemy wyrażenie

$$\hat{\beta}^{EN} = \arg\min_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \right\}.$$

Przez to, że mamy tutaj kwadratową karę z regresji grzbietowej, to funkcja jest wypukła i ma unikalne minimum. Jednkaże nie ma ono analitycznego rozwiązania i znalezienie $\hat{\beta}^{EN}$ sprowadza się do rozwiązania numerycznego.

Dziękuję za uwagę!