$$Y = \sum \beta_i X_i$$

Regresión Ridge.

El poblema de multicolinealidad surge cuando cuando la matriz de datos (matriz de diseño) $X = [X_1 \; X_2 \cdots X_p]$ no es de rango completo i.e. $(x_i \in \mathbb{R}^n)$

$$\dim \mathrm{Col}(X) = \mathrm{rk}(X) < \min(n,p)$$

$$\cancel{X_1} = 2X_2 + X_3$$

Intuitivamente hay una o mas columnas X_i que se pueden expresar como combinación lineal de otras columnas.

El modelo está sobreajusta. Hay variables que estan correlacionadas linealmente con otras.

El objetivo es elegir $S \subset \{X_1, \ldots, X_p\}$ de modo que S sea un conjunto linealmente independiente y

$$\mathrm{Col}(X) = \mathrm{Span}\, S.$$

Algunas fórmulas

$$VIF_j = \frac{1}{1 - R_j^2}$$

$$\kappa(X) = \frac{\sigma_{max}(X)}{\sigma_{min}(X)}$$ cuando la norma inducida es L^2.

VIFs

$$X_i = \sum_{j \neq i} \beta_j X_j$$ ← Modelo de Regresión

Sea el coeficiente R_i^2 de esta regresión.

Si $R_i^2 \approx 1$ entonces el modelo explica en mayor medida, la variabilidad de X_i

$$VIF_i = \frac{1}{1 - R_i^2}$$ si $R_i^2 \approx 1$ VIF_i es grande.

Número de Condición

Consideremos el sistema de ecuaciones

$Ax = b$. cambiamos b por e y tenemos el sistema $Ax = e$ $(e = b + \varepsilon)$

$$k(A) = \max_{e, b \neq 0} \left\{ \frac{\|\underbrace{A^{-1} b}_{sol\,1}\| / \|\underbrace{A^{-1} e}_{sol\,2}\|}{\|b\| / \|e\|} \right\}$$

El numero de condición de una matriz nos dice que tanto cambia la solución respecto a un cambio en el modelo

Si $k(A)$ es grande (> 30) entonces el problema es mal condicionado.

Def (Problema bien condicionado)

Un problema se dice bien condiciona si la solución varía de forma continua respecto al cambio en el modelo.

Para el problema de regresión debemos estimar β en el modelo

$$Y = X\beta + \varepsilon$$

El # de condición se calcula respecto a la matriz X. Para la norma L^2

$$K(X) = \sigma_{max}(X) / \sigma_{min}(X).$$

donde $\sigma_{max}(X)$ y $\sigma_{min}(X)$ son los valores singulares máximos y mínimos positivos.

Descomposición es Valores Singulares

Sea $X \in M_{m\times n}(\mathbb{R})$ existen $U \in M_{m\times r}(\mathbb{R})$, $V \in M_{n\times r}(\mathbb{R})$ matrices unitarias y $D = \text{diag}(d_1, \dots, d_r)$ con $d_i > 0$

$$\underset{m\times n}{X} = \underset{m\times r}{U}\ \underset{r\times r}{D}\ \underset{r\times n}{V^T}$$

donde r el rango de X.

Consideremos

$$X^T X = (V D U^T)(U D V^T)$$

$$= \underset{m\times r}{V}\ \underset{r\times r}{D^2}\ \underset{r\times m}{V^T}$$

y esta es la descomposición espectral de $X^T X$.

Numéricamente se usa en general M. Jacobi, Potencia Inversa.

Comentarios

1. La SVD es base de muchos métodos comp. y est. entre ellos PCA y Ridge.

2.. Teorema de Eckart - Young que justifica PCA como método de reducción de dimensionalidad.

3.. En un momento definimos la traza de Ridge.