

PCA muestral

Wednesday, April 6, 2022 9:33 AM

PCA muestral ##
Los vect y vals propios son alternos.
Le gusta reducir variabilidad de n mediciones sobre p variables.
Suponga q tiene los datos: x_1, \dots, x_n (n tiene una muestra p variable).
Sean μ, Σ la media y covarianza poblacional.
 \bar{x}, S \rightarrow media y covarianza muestral.
Le gusta. Un no correlacionados de datos muestrales q' max variabilidad son los CP (componentes principales).
 $\mu^T x_i$
 \rightarrow \hat{e}_i (vec propio de Σ)
Como la 1era CP maximiza $\hat{e}_1^T S \hat{e}_1$.
Los CP estaran dados por los val propios de Σ (muestral) u teoricos.
 $\hat{\lambda}_1 \rightarrow$ val propio asociado a \hat{e}_1 .
Ordenados: $\hat{\lambda}_1 > \dots > \hat{\lambda}_p > 0$ donde el val propio es el más y por tanto se puede hacer inferencia.
Varianza muestral de $\hat{y}_i = \hat{\lambda}_i$.
Cov muestral $\hat{y}_i, \hat{y}_j = 0 \rightarrow$ los CP son indep.
Var total muestral $= \sum_{i=1}^p \hat{\lambda}_i = \hat{\lambda}_1 + \dots + \hat{\lambda}_p$.
 $\hat{y}_i = \hat{e}_i^T S \hat{e}_i$
 $\frac{\hat{y}_i}{\sqrt{\hat{\lambda}_i}} = \frac{\hat{e}_i^T S \hat{e}_i}{\sqrt{\hat{\lambda}_i}}$ dete. std.
OBT:
1. Los CP se denotan $\hat{y}_1, \dots, \hat{y}_p$ indep de si se obtienen de S o con R.
2. Los CP pueden diferir si son de S o de R. (esto es cuando varían las variables).
3. Número apropiado de CPs (cada uno).
a) Ver la proporc de var total muestral capturada por los primeros k CP's.

b. Realizar gráfica
 \Rightarrow ver datos de cada punto para ver a partir de cual, agrega menos.
Tarea: Realizar entre análisis CP y RE (regresión conf).
Inferencia
 $\frac{\hat{\lambda}_1}{(1 + 3(n-1) \frac{\hat{\lambda}_1}{n})} \leq \lambda_1 \leq \frac{\hat{\lambda}_1}{(1 - 3(n-1) \frac{\hat{\lambda}_1}{n})}$