

Analyse en composante principales

Introduction

Véronique Tremblay

Exemple

Ville	Précipitations		Température			
	Août		Août		Septembre	
	Neige (cm)	Pluie (mm)	°C	°F	°C	°F
a	0	104	25	77	22	72
b	0	97	23	73	21	70
c	0	96	27	81	22	72
d	0	107	30	86	28	82
e	0	82	33	91	31	88

L'idée de l'ACP

On cherche une combinaison linéaire des variables qui maximise la variance.

$$Y = a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + a_5X_5 + a_6X_6$$

Ville	Précipitations		Température			
	Août		Août		Septembre	
	Neige	Pluie	°C	°F	°C	°F
	X1	X2	X3	X4	X5	X6
a	0	104	25	77	22	72
b	0	97	23	73	21	70
c	0	96	27	81	22	72
d	0	107	30	86	28	82
e	0	82	33	91	31	88

$$Y = a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + a_5X_5 + a_6X_6$$

En étudiant les corrélations?

	Précipitations		Température			
	Août		Août		Septembre	
	Neige	Pluie	°C	°F	°C	°F
	X1	X2	X3	X4	X5	X6
X1	1	NA	NA	NA	NA	NA
X2	NA	1.00	-0.48	-0.47	-0.45	-0.48
X3	NA	-0.48	1.00	1.00	0.96	0.96
X4	NA	-0.47	1.00	1.00	0.95	0.95
X5	NA	-0.45	0.96	0.95	1.00	1.00
X6	NA	-0.48	0.96	0.95	1.00	1.00

Pourquoi fait-on l'ACP?

- Explorer un jeu de données de grande dimension
- Faciliter la construction de modèles