Vorgehen

Hinweise zu: **Aufgabe 1** **Aufgabe 2**
1a 2b

Hinweise zu Aufgabe 1

a. Die neuen Koordinatenachsen entsprechen gerade den gelernten Eigenvektoren. Das hängt damit zusammen, daß es sich um eine orthonormale Transformation, also eine Rotation im Eingaberaum har die die Eigenvektoren auf die Koordinatenachsen des ursprünglichen Koordinatensystems rotiert.

nach oben

Hinweise zu Aufgabe 2

b. Um die Eigenvektoren zu Eigenwert 0 zu bestimmen, können Sie da Gram-Schmidt Orthogonalisierungsverfahren verwenden. Nachdem d Sangeralgorithmus alle Eigenvektoren $\underline{w}_1, ..., \underline{w}_m$ zu Eigenwerten >0 berechnet hat, können Sie für die Eigenvektoren , die zu Eigenwert 0 gehören, zufällige Vektoren wählen und diese dann für $m + 1 \leq j \leq r$

$$\begin{array}{lcl} \underline{w}_j & \leftarrow & \underline{w}_j - \sum_{i=1}^{j-1} \underline{w}_i^T \underline{w}_j \cdot \underline{w}_i \\ \underline{w}_j & \leftarrow & \underline{w}_j \cdot \frac{1}{||\underline{w}_j||_2} \end{array}$$

orthogonalisieren. Da die so gebildeten Vektoren senkrecht auf dem F stehen, der durch die Eigenvektoren zu Eigenwerten >0 aufgespannt wird, dieser Raum aber die gesamten Datenpunkte enthält, sind die so orthogonalisierten Vektoren Eigenvektoren zu Eigenwert 0. Für jedes es nämlich Zahlen $a_1, ..., a_m \in \mathbb{R}$, so daß für gilt:

$$
\begin{array}{rrcl}
 & \underline{x} & = & \sum_{i=1}^{m} a_i \cdot \underline{w}_i \\
\Longrightarrow & C_{xx}\underline{w}_j & = & \left\langle \underline{x}\underline{x}^T \right\rangle \underline{w}_j \\
 & & = & \left\langle \underline{x}\underline{x}^T \underline{w}_j \right\rangle \\
 & & = & \left\langle \underline{x} \cdot \sum_{i=1}^{m} a_i \cdot \underbrace{\underline{w}_i^T \underline{w}_j}_{=0} \right\rangle \\
 & & = & 0 \\
\Longrightarrow & C_{xx}\underline{w}_j & = & 0 \cdot \underline{w}_j
\end{array}
$$

Insbesondere sieht man anhand der Rechnung, daß für alle $\underline{x}$ die Beziehung $\underline{x}^T \underline{w}_j = 0$ gilt. Diesen Zusammenhang können Sie verwenden, um die Spalten des `card1`-Datensatzes zu bestimmen, die keine weitere Information zu den $\underline{x}$ beitragen.

(Zur Erinnerung: Eine Sequenz von Werten enthält keine zusätzliche (neue) Information über eine andere Sequenz, wenn die Werte der einen stets aus den Werten der anderen berechnet werden können)

nach oben