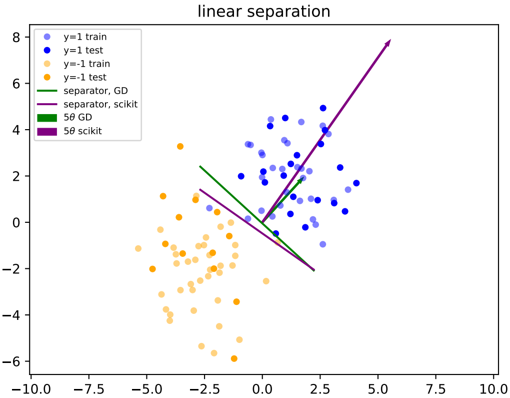
Fondamentaux théoriques du machine learning



Overview of lecture 4

PCA

K-means clustering

First principal component

We look for w, $||w|| = 1$ such that

$$\sum_{i=1}^{n} \left(w^T x_i\right)^2 \tag{1}$$

is maximal.

Proposition

w is the eigenvector of $X^T X$ with largest eigenvalue λ_{max}.

First principal component

We look for w, $||w|| = 1$ such that

$$\sum_{i=1}^{n} \left(w^T x_i\right)^2 \tag{2}$$

is maximal.

Proposition

w is the eigenvector of $X^T X$ with largest eigenvalue λ_{max}.

Exercice 1 : Show the proposition.

First principal component

$$\sum_{i=1}^{n}(w^T x_i)^2 = ||Xw||^2$$
$$= \langle Xw, Xw \rangle$$
$$= \langle (X^T X)w, w \rangle$$

This quantity is always smaller that λ_{max}, and it attained for an eigenvector in the eigenspace with norm 1, since we impose that $||w|| = 1$.

Minimization

Exercice 2 : **Gradient :**
Compute the gradient of $J(\Omega, z)$ with respect to Ω and deduce the minimizer Ω^*.

- z is fixed
- we can see Ω has a vector of $\mathbb{R}^{Kd}$.

It is sufficient to compute the gradient with respect to all the ω_k separately.

Minimization

We compute the gradient of J with respect to ω_{k_0}.

$$\begin{aligned}
\nabla_{\omega_{k_0}} J &= \nabla_{\omega_{k_0}} \sum_{i=1}^{n} \sum_{k=1}^{K} z_i^k ||x_i - \omega_k||^2 \\
&= \sum_{i=1}^{n} \nabla_{\omega_{k_0}} \sum_{k=1}^{K} z_i^k ||x_i - \omega_k||^2 \\
&= \sum_{i=1}^{n} z_i^{k_0} \nabla_{\omega_{k_0}} ||x_i - \omega_{k_0}||^2
\end{aligned} \tag{3}$$

We need to compute

$$\nabla_{\omega_{k_0}} ||x_i - \omega_{k_0}||^2 \tag{4}$$

Let

$$u = \begin{cases} \mathbb{R}^d \to \mathbb{R} \\ y \mapsto ||y||^2 \end{cases}$$

$$v_i = \begin{cases} \mathbb{R}^d \to \mathbb{R} \\ \omega \mapsto x_i - \omega \end{cases}$$

$$w_i = \begin{cases} \mathbb{R}^d \to \mathbb{R} \\ \omega \mapsto ||x_i - \omega||^2 \end{cases}$$

We have

$$w_i = u \circ v_i \tag{5}$$

Differentials

We already now that

$$\forall (y,h) \in (\mathbb{R}^d)^2, du_y(h) = \langle 2y, h\rangle \tag{6}$$

and that

$$\forall (\omega,h) \in (\mathbb{R}^d)^2, d(v_i)_\omega(h) = -h \tag{7}$$

By composition,

$$\nabla_{\omega_{k_0}} ||x_i - \omega_{k_0}||^2 = -2(x_i - \omega_{k_0}) \tag{8}$$

Gradient cancellation

This gradient cancels if

$$\begin{aligned}
\nabla_{\omega_{k_0}} J &= \sum_{i=1}^{n} z_i^{k_0} \nabla_{\omega_{k_0}} ||x_i - \omega_{k_0}||^2 \\
&= -\sum_{i=1}^{n} z_i^{k_0} 2(x_i - \omega_{k_0}) \\
&= 0
\end{aligned} \tag{9}$$

Gradient cancellation

This gradient cancels if

$$2\omega_{k_0}\sum_{i=1}^{n} z_i^{k_0} = 2\sum_{i=1}^{n} z_i^{k_0} x_i \tag{10}$$

or equivalently

$$\omega_{k_0} = \frac{\sum_{i=1}^{n} z_i^{k_0} x_i}{\sum_{i=1}^{n} z_i^{k_0}} \tag{11}$$

Hence, the minimizer $w_{k_0}^*$ is the average of its cluster.