Otros Algoritmos con Kernels

Fernando Lozano

Universidad de los Andes

31 de octubre de 2022

SVMs

$$
\begin{aligned}
\text{mín} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{n}\sum_{i=1}^{n}\zeta_i \\
\text{sujeto a} \quad & y_i\left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b\right) - 1 + \zeta_i \geq 0 \quad i = 1, \ldots, n \\
& \zeta_i \geq 0
\end{aligned}
$$

SVMs

$$
\begin{aligned}
\text{mín} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{n}\sum_{i=1}^{n}\zeta_i \\
\text{sujeto a} \quad & y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i \geq 0 \quad i = 1, \ldots, n \\
& \zeta_i \geq 0
\end{aligned}
$$

$$
\begin{aligned}
\text{máx} \quad & \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n}\alpha_i\alpha_j y_i y_j \left\langle \mathbf{x}_i, \mathbf{x}_j\right\rangle \\
\text{sujeto a} \quad & \sum_{i=1}^{n}\alpha_i y_i = 0 \\
& 0 \leq \alpha_i \leq \frac{C}{n}
\end{aligned}
$$

El truco del Kernel

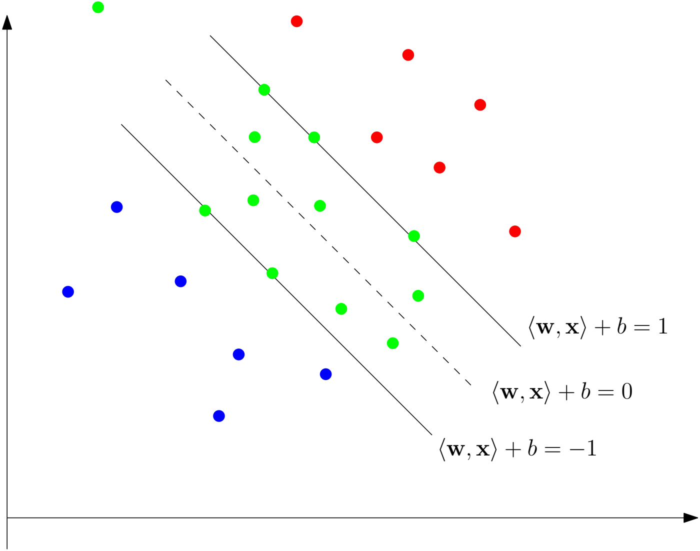
$$
\begin{aligned}
\text{máx} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) \\
\text{sujeto a} \quad & \sum_{i=1}^{n} \alpha_i y_i = 0 \\
& 0 \leq \alpha_i \leq \frac{C}{n}
\end{aligned}
$$

El truco del Kernel

$$\begin{aligned}
\text{máx} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) \\
\text{sujeto a} \quad & \sum_{i=1}^{n} \alpha_i y_i = 0 \\
& 0 \leq \alpha_i \leq \frac{C}{n}
\end{aligned}$$

- Solución: $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i) \boldsymbol{\phi}(\mathbf{x}_i)$
- Evaluación:

$$\begin{aligned}
\operatorname{sign}\left(\langle w, \boldsymbol{\phi}(\mathbf{x}) \rangle_{\mathcal{H}} + b\right) &= \operatorname{sign}\left(\sum_{i=1}^{n} (\alpha_i y_i) \left\langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}) \right\rangle_{\mathcal{H}} + b \right) \\
&= \operatorname{sign}\left(\sum_{i=1}^{n} (\alpha_i y_i) k(\mathbf{x}_i, \mathbf{x}) + b \right)
\end{aligned}$$


$\langle \mathbf{w}, \mathbf{x} \rangle + b = 1$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = -1$

ν-SVM

- C: compromiso entre márgen y error empírico.

ν-SVM

- C: compromiso entre márgen y error empírico.
- No es fácil de determinar a priori.

ν-SVM

- C: compromiso entre márgen y error empírico.
- No es fácil de determinar a priori.
- No tiene significado intuitivo.

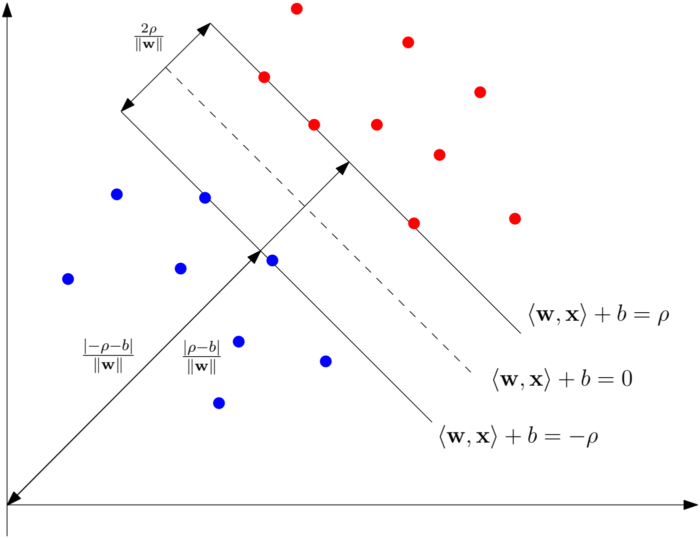
ν-SVM

- C: compromiso entre márgen y error empírico.
- No es fácil de determinar a priori.
- No tiene significado intuitivo.
- ν-SVM reemplaza C por un nuevo parámetro ν que controla el número de errores y el número de vectores de soporte:

ν-SVM

- C: compromiso entre márgen y error empírico.
- No es fácil de determinar a priori.
- No tiene significado intuitivo.
- ν-SVM reemplaza C por un nuevo parámetro ν que controla el número de errores y el número de vectores de soporte:

$$\begin{aligned}
\min_{\mathbf{w},b,\zeta_i,\rho} \quad & \frac{1}{2}\|\mathbf{w}\|^2 - \nu\rho + \frac{1}{n}\sum_{i=1}^{n}\zeta_i \\
\text{sujeto a} \quad & y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) + \zeta_i - \rho \geq 0 \quad i = 1, \ldots, n \\
& \zeta_i \geq 0, \rho \geq 0
\end{aligned}$$



$\frac{2\rho}{\|\mathbf{w}\|}$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = \rho$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = -\rho$
$\frac{|-\rho-b|}{\|\mathbf{w}\|}$
$\frac{|\rho-b|}{\|\mathbf{w}\|}$

- Lagrangiano ($\delta \geq 0, \boldsymbol{\zeta} \geq 0$):

$$L(\mathbf{w}, b, \boldsymbol{\zeta}, \rho, \boldsymbol{\alpha}, \boldsymbol{\mu}, \delta) = \frac{1}{2}\|\mathbf{w}\|^2 - \nu\rho + \frac{1}{n}\sum_{i=1}^{n}\zeta_i - \sum_{i=1}^{n}\alpha_i\left(y_i\left(\langle\mathbf{w}, \mathbf{x}_i\rangle + b\right) - \rho + \zeta_i\right) - \sum_{i=1}^{n}\mu_i\zeta_i - \delta\rho$$

- Lagrangiano ($\delta \geq 0, \boldsymbol{\zeta} \geq 0$):

$$L(\mathbf{w}, b, \boldsymbol{\zeta}, \rho, \boldsymbol{\alpha}, \boldsymbol{\mu}, \delta) = \frac{1}{2}\|\mathbf{w}\|^2 - \nu\rho + \frac{1}{n}\sum_{i=1}^{n}\zeta_i - \sum_{i=1}^{n}\alpha_i\left(y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - \rho + \zeta_i\right) - \sum_{i=1}^{n}\mu_i\zeta_i - \delta\rho$$

- Derivando con respecto a $\mathbf{w}, b, \boldsymbol{\zeta}, \rho$, e igualando a cero:

$$\mathbf{w} = \sum_{i=1}^{n}\alpha_i y_i \mathbf{x}_i$$

$$\alpha_i + \mu_i = 1/n$$

$$\sum_{i=1}^{n}\alpha_i y_i = 0$$

$$\sum_{i=1}^{n}\alpha_i - \delta = \nu$$

- Problema dual:

$$
\begin{aligned}
\text{máx} \quad & -\frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n}\alpha_i\alpha_j y_i y_j k(\mathbf{x}_i,\mathbf{x}_j) \\
\text{sujeto a} \quad & 0 \le \alpha_i \le \frac{1}{n} \\
& \sum_{i=1}^{n}\alpha_i y_i = 0 \\
& \sum_{i=1}^{n}\alpha_i \ge \nu
\end{aligned}
$$

Propiedades

Propiedades

- Error de márgen: datos que están mal clasificados o tienen márgen menos que ρ

Propiedades

- Error de márgen: datos que están mal clasificados o tienen márgen menos que ρ (es decir $\zeta_i > 0$).

Propiedades

- Error de márgen: datos que están mal clasificados o tienen márgen menos que ρ (es decir $\zeta_i > 0$).
- Si en el óptimo $\rho > 0$, ν es una cota superior de el porcentaje de errores de márgen.

Propiedades

- Error de márgen: datos que están mal clasificados o tienen márgen menos que ρ (es decir $\zeta_i > 0$).
- Si en el óptimo $\rho > 0$, ν es una cota superior de el porcentaje de errores de márgen.
  - Por KKT, $\rho > 0 \Rightarrow \delta = 0$, luego $\sum_{i=1}^{n} \alpha_i = \nu$,

Propiedades

- Error de márgen: datos que están mal clasificados o tienen márgen menos que ρ (es decir $\zeta_i > 0$).
- Si en el óptimo $\rho > 0$, ν es una cota superior de el porcentaje de errores de márgen.
  - Por KKT, $\rho > 0 \Rightarrow \delta = 0$, luego $\sum_{i=1}^{n} \alpha_i = \nu$, y $\zeta_i > 0 \Rightarrow \alpha_i = 1/n$

Propiedades

- Error de márgen: datos que están mal clasificados o tienen márgen menos que ρ (es decir $\zeta_i > 0$).
- Si en el óptimo $\rho > 0$, ν es una cota superior de el porcentaje de errores de márgen.
  - Por KKT, $\rho > 0 \Rightarrow \delta = 0$, luego $\sum_{i=1}^{n} \alpha_i = \nu$, y $\zeta_i > 0 \Rightarrow \alpha_i = 1/n$
- Si en el óptimo $\rho > 0$, ν es una cota inferior de la fracción de vectores de soporte.

Propiedades

- Error de márgen: datos que están mal clasificados o tienen márgen menos que ρ (es decir $\zeta_i > 0$).
- Si en el óptimo $\rho > 0$, ν es una cota superior de el porcentaje de errores de márgen.
  - Por KKT, $\rho > 0 \Rightarrow \delta = 0$, luego $\sum_{i=1}^{n} \alpha_i = \nu$, y $\zeta_i > 0 \Rightarrow \alpha_i = 1/n$
- Si en el óptimo $\rho > 0$, ν es una cota inferior de la fracción de vectores de soporte.
  - $\sum_{i=1}^{n} \alpha_i \geq \nu$,

Propiedades

- Error de márgen: datos que están mal clasificados o tienen márgen menos que ρ (es decir $\zeta_i > 0$).
- Si en el óptimo $\rho > 0$, ν es una cota superior de el porcentaje de errores de márgen.
  - Por KKT, $\rho > 0 \Rightarrow \delta = 0$, luego $\sum_{i=1}^{n} \alpha_i = \nu$, y $\zeta_i > 0 \Rightarrow \alpha_i = 1/n$
- Si en el óptimo $\rho > 0$, ν es una cota inferior de la fracción de vectores de soporte.
  - $\sum_{i=1}^{n} \alpha_i \geq \nu$, un vector de soporte tiene $\alpha_i \leq 1/n$.

Propiedades

- Error de márgen: datos que están mal clasificados o tienen márgen menos que ρ (es decir $\zeta_i > 0$).
- Si en el óptimo $\rho > 0$, ν es una cota superior de el porcentaje de errores de márgen.
  - Por KKT, $\rho > 0 \Rightarrow \delta = 0$, luego $\sum_{i=1}^{n} \alpha_i = \nu$, y $\zeta_i > 0 \Rightarrow \alpha_i = 1/n$
- Si en el óptimo $\rho > 0$, ν es una cota inferior de la fracción de vectores de soporte.
  - $\sum_{i=1}^{n} \alpha_i \geq \nu$, un vector de soporte tiene $\alpha_i \leq 1/n$.
- Bajo ciertas condiciones suaves, asimptóticamente con probabilidad 1, ν es igual a la fracción de errores de márgen y a la fracción de vectores de soporte.

Propiedades

- Error de márgen: datos que están mal clasificados o tienen márgen menos que ρ (es decir $\zeta_i > 0$).
- Si en el óptimo $\rho > 0$, ν es una cota superior de el porcentaje de errores de márgen.
    - Por KKT, $\rho > 0 \Rightarrow \delta = 0$, luego $\sum_{i=1}^{n} \alpha_i = \nu$, y $\zeta_i > 0 \Rightarrow \alpha_i = 1/n$
- Si en el óptimo $\rho > 0$, ν es una cota inferior de la fracción de vectores de soporte.
    - $\sum_{i=1}^{n} \alpha_i \geq \nu$, un vector de soporte tiene $\alpha_i \leq 1/n$.
- Bajo ciertas condiciones suaves, asimptóticamente con probabilidad 1, ν es igual a la fracción de errores de márgen y a la fracción de vectores de soporte.
- La solución de ν-SVM con $\rho > 0$, es la misma solución que se obtiene para el SVM con $C = 1/\rho$.

Ejemplo

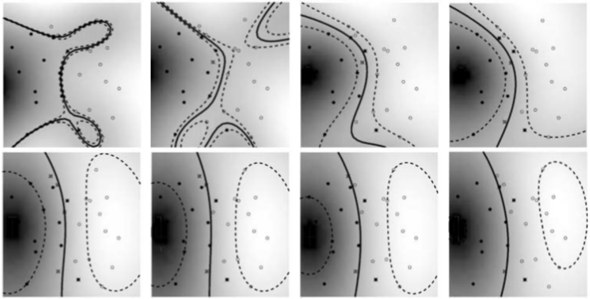


Figure 7.9 Toy problem (task: separate circles from disks) solved using ν-SV classification, with parameter values ranging from $\nu = 0.1$ (top left) to $\nu = 0.8$ (bottom right). The larger we make ν, the more points are allowed to lie inside the margin (depicted by dotted lines). Results are shown for a Gaussian kernel, $k(x, x') = \exp(-\|x - x'\|^2)$.

ν	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
fraction of errors	0.00	0.07	0.25	0.32	0.39	0.50	0.61	0.71
fraction of SVs	0.29	0.36	0.43	0.46	0.57	0.68	0.79	0.86
margin $\rho/\|\mathbf{w}\|$	0.005	0.018	0.115	0.156	0.364	0.419	0.461	0.546

Support Vector Regression

- En el problema de regresión queremos aprender una función real $y = f(\mathbf{x})$ a partir de observaciones $\{\mathbf{x}_i, y_i\}_{i=1}^{n}$.

Support Vector Regression

- En el problema de regresión queremos aprender una función real $y = f(\mathbf{x})$ a partir de observaciones $\{\mathbf{x}_i, y_i\}_{i=1}^n$.
- Criterio de error:

$$R(f) = \mathbf{E}_{z\sim\mathcal{D}}\left[c(f, \mathbf{x}, y)\right]$$

Support Vector Regression

- En el problema de regresión queremos aprender una función real $y = f(\mathbf{x})$ a partir de observaciones $\{\mathbf{x}_i, y_i\}_{i=1}^n$.
- Criterio de error:

$$R(f) = \mathbf{E}_{z \sim \mathcal{D}} \left[c(f, \mathbf{x}, y) \right]$$

- Por ejemplo:

$$R(f) = \mathbf{E}_{z \sim \mathcal{D}} \left[(f(\mathbf{x}) - y)^2 \right]$$

Support Vector Regression

- En el problema de regresión queremos aprender una función real $y = f(\mathbf{x})$ a partir de observaciones $\{\mathbf{x}_i, y_i\}_{i=1}^n$.
- Criterio de error:
$$R(f) = \mathbf{E}_{z\sim\mathcal{D}}\left[c(f, \mathbf{x}, y)\right]$$
- Por ejemplo:
$$R(f) = \mathbf{E}_{z\sim\mathcal{D}}\left[(f(\mathbf{x}) - y)^2\right]$$
- Modelo es una función afín en espacio de características:

Support Vector Regression

- En el problema de regresión queremos aprender una función real $y = f(\mathbf{x})$ a partir de observaciones $\{\mathbf{x}_i, y_i\}_{i=1}^n$.
- Criterio de error:

$$R(f) = \mathbf{E}_{z\sim\mathcal{D}}\left[c(f, \mathbf{x}, y)\right]$$

- Por ejemplo:

$$R(f) = \mathbf{E}_{z\sim\mathcal{D}}\left[(f(\mathbf{x}) - y)^2\right]$$

- Modelo es una función afín en espacio de características:

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b$$

Support Vector Regression

- En el problema de regresión queremos aprender una función real $y = f(\mathbf{x})$ a partir de observaciones $\{\mathbf{x}_i, y_i\}_{i=1}^n$.
- Criterio de error:

$$R(f) = \mathbf{E}_{z\sim\mathcal{D}}\left[c(f, \mathbf{x}, y)\right]$$

- Por ejemplo:

$$R(f) = \mathbf{E}_{z\sim\mathcal{D}}\left[(f(\mathbf{x}) - y)^2\right]$$

- Modelo es una función afín en espacio de características:

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b$$

- Idea: minimizar función de error:

$$\frac{1}{2}\|\mathbf{w}\|^2 + CR_n(f)$$

donde $R_n(f)$ mide error en los datos.

Support Vector Regression

- En el problema de regresión queremos aprender una función real $y = f(\mathbf{x})$ a partir de observaciones $\{\mathbf{x}_i, y_i\}_{i=1}^n$.
- Criterio de error:

$$R(f) = \mathbf{E}_{z\sim\mathcal{D}}\left[c(f, \mathbf{x}, y)\right]$$

- Por ejemplo:

$$R(f) = \mathbf{E}_{z\sim\mathcal{D}}\left[(f(\mathbf{x}) - y)^2\right]$$

- Modelo es una función afín en espacio de características:

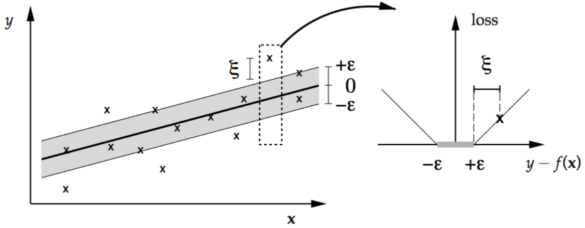
$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b$$

- Idea: minimizar función de error:

$$\frac{1}{2}\|\mathbf{w}\|^2 + CR_n(f)$$

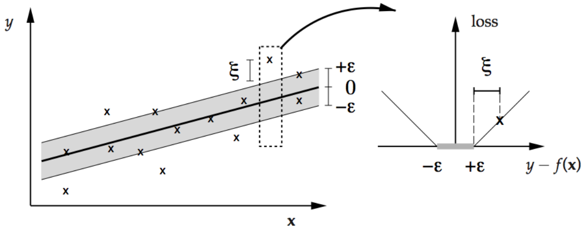
donde $R_n(f)$ mide error en los datos.

Función de pérdida ϵ-insensitiva



$$|y - f(\mathbf{x})|_\epsilon = \text{máx}\,\{0, |y - f(\mathbf{x})| - \epsilon\}$$

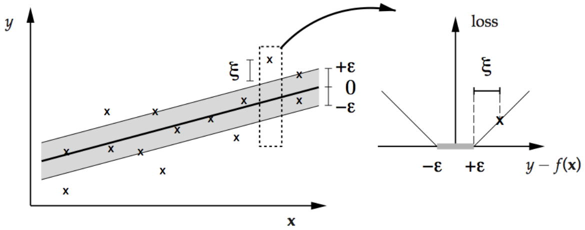
Función de pérdida ϵ-insensitiva



$$|y - f(\mathbf{x})|_\epsilon = \max\left\{0, |y - f(\mathbf{x})| - \epsilon\right\}$$

- Errores dentro del tubo no contribuyen al error.

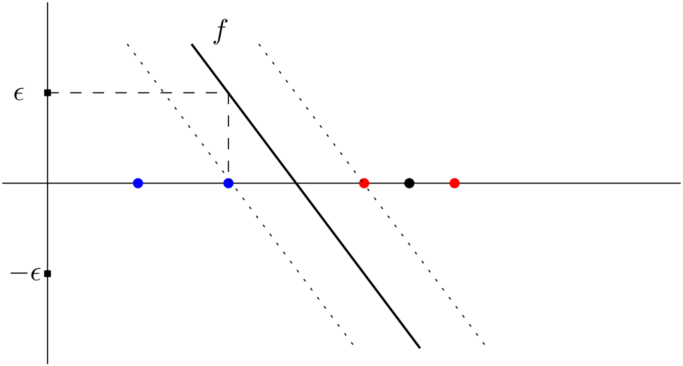
Función de pérdida ϵ-insensitiva



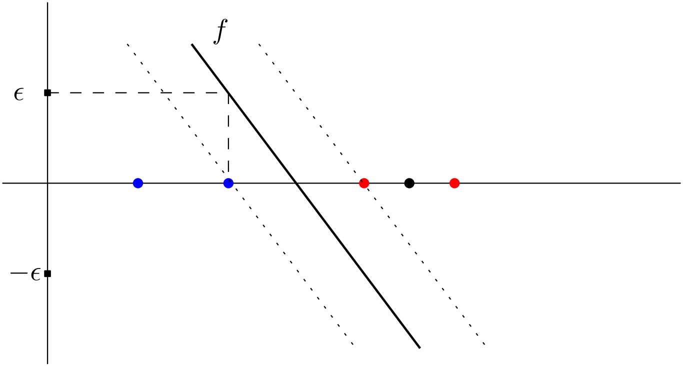
$$|y - f(\mathbf{x})|_{\epsilon} = \text{máx}\left\{0, |y - f(\mathbf{x})| - \epsilon\right\}$$

- Errores dentro del tubo no contribuyen al error.
- En SVMs datos con margen no contribuyen al error.

Relación con SVMs

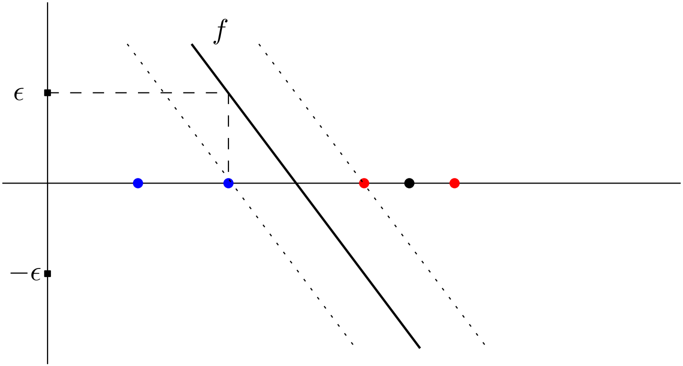


Relación con SVMs



- SVR: $\|\mathbf{w}\| \lll$ corresponde a función plana.

Relación con SVMs



- SVR: $\|\mathbf{w}\| \lll$ corresponde a función plana.
- SVM: $\|\mathbf{w}\| \lll$ corresponde a márgen grande.

ϵ–SVR

$$
\begin{aligned}
\text{mín} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{n}\sum_{i=1}^{n}(\zeta_i + \zeta_i^*) \\
\text{sujeto a} \quad & (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - y_i \leq \epsilon + \zeta_i \\
& y_i - (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \leq \epsilon + \zeta_i^* \\
& \zeta_i, \zeta_i^* \geq 0
\end{aligned}
$$

Problema dual

Problema dual

- Lagrangiano $(\alpha, \alpha^*, \eta, \eta^* \geq 0)$:

$$L = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{n}\sum_{i=1}^{n}(\zeta_i + \zeta_i^*) - \sum_{i=1}^{n}(\eta_i\zeta_i + \eta_i^*\zeta_i^*)$$

$$-\sum_{i=1}^{n}\alpha_i(\epsilon + \zeta_i + y_i - \langle\mathbf{w}, \mathbf{x}_i\rangle - b) - \sum_{i=1}^{n}\alpha_i^*(\epsilon + \zeta_i^* - y_i + \langle\mathbf{w}, \mathbf{x}_i\rangle - b)$$

Problema dual

- Lagrangiano ($\alpha, \alpha^*, \eta, \eta^* \geq 0$):

$$L = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{n}\sum_{i=1}^{n}(\zeta_i + \zeta_i^*) - \sum_{i=1}^{n}(\eta_i\zeta_i + \eta_i^*\zeta_i^*)$$
$$-\sum_{i=1}^{n}\alpha_i(\epsilon + \zeta_i + y_i - \langle \mathbf{w}, \mathbf{x}_i\rangle - b) - \sum_{i=1}^{n}\alpha_i^*(\epsilon + \zeta_i^* - y_i + \langle \mathbf{w}, \mathbf{x}_i\rangle - b)$$

- Derivando con respecto a las variables primales e igualando a cero:

Problema dual

- Lagrangiano $(\alpha, \alpha^*, \eta, \eta^* \geq 0)$:

$$L = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{n}\sum_{i=1}^{n}(\zeta_i + \zeta_i^*) - \sum_{i=1}^{n}(\eta_i\zeta_i + \eta_i^*\zeta_i^*)$$

$$-\sum_{i=1}^{n}\alpha_i(\epsilon + \zeta_i + y_i - \langle \mathbf{w}, \mathbf{x}_i\rangle - b) - \sum_{i=1}^{n}\alpha_i^*(\epsilon + \zeta_i^* - y_i + \langle \mathbf{w}, \mathbf{x}_i\rangle - b)$$

- Derivando con respecto a las variables primales e igualando a cero:

$$\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0$$

Problema dual

- Lagrangiano ($\alpha, \alpha^*, \eta, \eta^* \geq 0$):

$$L = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{n}\sum_{i=1}^{n}(\zeta_i + \zeta_i^*) - \sum_{i=1}^{n}(\eta_i\zeta_i + \eta_i^*\zeta_i^*)$$

$$-\sum_{i=1}^{n}\alpha_i(\epsilon + \zeta_i + y_i - \langle \mathbf{w}, \mathbf{x}_i\rangle - b) - \sum_{i=1}^{n}\alpha_i^*(\epsilon + \zeta_i^* - y_i + \langle \mathbf{w}, \mathbf{x}_i\rangle - b)$$

- Derivando con respecto a las variables primales e igualando a cero:

$$\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0$$

$$\mathbf{w} = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\mathbf{x}_i$$

Problema dual

- Lagrangiano ($\alpha, \alpha^*, \eta, \eta^* \geq 0$):

$$L = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{n}\sum_{i=1}^{n}(\zeta_i + \zeta_i^*) - \sum_{i=1}^{n}(\eta_i\zeta_i + \eta_i^*\zeta_i^*)$$
$$-\sum_{i=1}^{n}\alpha_i(\epsilon + \zeta_i + y_i - \langle \mathbf{w}, \mathbf{x}_i\rangle - b) - \sum_{i=1}^{n}\alpha_i^*(\epsilon + \zeta_i^* - y_i + \langle \mathbf{w}, \mathbf{x}_i\rangle - b)$$

- Derivando con respecto a las variables primales e igualando a cero:

$$\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0$$

$$\mathbf{w} = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\mathbf{x}_i$$

$$\alpha_i - \eta_i = \alpha_i^* - \eta_i^* = 0$$

Problema Dual

$$
\begin{aligned}
\text{máx} \quad & -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)\left\langle \mathbf{x}_i, \mathbf{x}_j \right\rangle - \epsilon \sum_{i=1}^{n}(\alpha_i^* + \alpha_i) \\
& + \sum_{i=1}^{n} y_i(\alpha_i^* - \alpha_i) \\
\text{sujeto a} \quad & \sum_{i=1}^{n}(\alpha_i^* - \alpha_i) = 0 \\
& \alpha_i, \alpha_i^* \in [0, C/n]
\end{aligned}
$$

Problema Dual

$$
\begin{aligned}
\text{máx} \quad & -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)\,{\color{red}\langle \mathbf{x}_i, \mathbf{x}_j \rangle} - \epsilon\sum_{i=1}^{n}(\alpha_i^* + \alpha_i) \\
& + \sum_{i=1}^{n} y_i(\alpha_i^* - \alpha_i) \\
\text{sujeto a} \quad & \sum_{i=1}^{n}(\alpha_i^* - \alpha_i) = 0 \\
& \alpha_i, \alpha_i^* \in [0, C/n]
\end{aligned}
$$

- Datos en términos de productos puntos

Problema Dual

$$
\begin{aligned}
\text{máx} \quad & -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)\,{\color{red}\langle \mathbf{x}_i, \mathbf{x}_j \rangle} - \epsilon\sum_{i=1}^{n}(\alpha_i^* + \alpha_i) \\
& + \sum_{i=1}^{n} y_i(\alpha_i^* - \alpha_i) \\
\text{sujeto a} \quad & \sum_{i=1}^{n}(\alpha_i^* - \alpha_i) = 0 \\
& \alpha_i, \alpha_i^* \in [0, C/n]
\end{aligned}
$$

- Datos en términos de productos puntos $\Rightarrow$ truco del kernel.

KKT

1. $\alpha_i(\epsilon + \zeta_i - y_i + \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0$

KKT

1. $\alpha_i(\epsilon + \zeta_i - y_i + \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0$
2. $\alpha_i^*(\epsilon + \zeta_i^* + y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0$

KKT

1. $\alpha_i(\epsilon + \zeta_i - y_i + \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0$
2. $\alpha_i^*(\epsilon + \zeta_i^* + y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0$
3. $\left(\frac{C}{n} - \alpha_i\right) \zeta_i = 0$

KKT

1. $\alpha_i(\epsilon + \zeta_i - y_i + \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0$
2. $\alpha_i^*(\epsilon + \zeta_i^* + y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0$
3. $\left(\frac{C}{n} - \alpha_i\right) \zeta_i = 0$
4. $\left(\frac{C}{n} - \alpha_i^*\right) \zeta_i^* = 0$

KKT

1. $\alpha_i(\epsilon + \zeta_i - y_i + \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0$
2. $\alpha_i^*(\epsilon + \zeta_i^* + y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0$
3. $\left(\frac{C}{n} - \alpha_i\right) \zeta_i = 0$
4. $\left(\frac{C}{n} - \alpha_i^*\right) \zeta_i^* = 0$

KKT

1. $\alpha_i(\epsilon + \zeta_i - y_i + \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0$
2. $\alpha_i^*(\epsilon + \zeta_i^* + y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0$
3. $\left(\frac{C}{n} - \alpha_i\right) \zeta_i = 0$
4. $\left(\frac{C}{n} - \alpha_i^*\right) \zeta_i^* = 0$

Consecuencias:

KKT

1. $\alpha_i(\epsilon + \zeta_i - y_i + \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0$
2. $\alpha_i^*(\epsilon + \zeta_i^* + y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0$
3. $\left(\frac{C}{n} - \alpha_i\right) \zeta_i = 0$
4. $\left(\frac{C}{n} - \alpha_i^*\right) \zeta_i^* = 0$

Consecuencias:

- Sólo datos para los cuales $\alpha_i = \frac{C}{n}$ o $\alpha_i^* = \frac{C}{n}$ pueden estar fuera del tubo ($\zeta_i > 0$).

KKT

1. $\alpha_i(\epsilon + \zeta_i - y_i + \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0$
2. $\alpha_i^*(\epsilon + \zeta_i^* + y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0$
3. $\left(\frac{C}{n} - \alpha_i\right) \zeta_i = 0$
4. $\left(\frac{C}{n} - \alpha_i^*\right) \zeta_i^* = 0$

Consecuencias:

- Sólo datos para los cuales $\alpha_i = \frac{C}{n}$ o $\alpha_i^* = \frac{C}{n}$ pueden estar fuera del tubo ($\zeta_i > 0$).
- $\alpha_i \alpha_i^* = 0$

KKT

1. $\alpha_i(\epsilon + \zeta_i - y_i + \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0$
2. $\alpha_i^*(\epsilon + \zeta_i^* + y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0$
3. $\left(\frac{C}{n} - \alpha_i\right) \zeta_i = 0$
4. $\left(\frac{C}{n} - \alpha_i^*\right) \zeta_i^* = 0$

Consecuencias:

- Sólo datos para los cuales $\alpha_i = \frac{C}{n}$ o $\alpha_i^* = \frac{C}{n}$ pueden estar fuera del tubo ($\zeta_i > 0$).
- $\alpha_i \alpha_i^* = 0$
- Para $\alpha_i \in (0, \frac{C}{n})$ (o $\alpha_i^* \in (0, \frac{C}{n})$), podemos calcular:

$$b = y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - \epsilon$$

KKT

1. $\alpha_i(\epsilon + \zeta_i - y_i + \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0$
2. $\alpha_i^*(\epsilon + \zeta_i^* + y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0$
3. $\left(\frac{C}{n} - \alpha_i\right) \zeta_i = 0$
4. $\left(\frac{C}{n} - \alpha_i^*\right) \zeta_i^* = 0$

Consecuencias:

- Sólo datos para los cuales $\alpha_i = \frac{C}{n}$ o $\alpha_i^* = \frac{C}{n}$ pueden estar fuera del tubo ($\zeta_i > 0$).
- $\alpha_i \alpha_i^* = 0$
- Para $\alpha_i \in (0, \frac{C}{n})$ (o $\alpha_i^* \in (0, \frac{C}{n})$), podemos calcular:

$$b = y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - \epsilon$$

- Vectores de soporte son los vectores fuera del tubo.

ν-SVR

$$
\begin{aligned}
\text{mín} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\left(\nu\epsilon + \frac{1}{n}\sum_{i=1}^{n}(\zeta_i + \zeta_i^*)\right) \\
\text{sujeto a} \quad & (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - y_i \leq \epsilon + \zeta_i \\
& y_i - (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \leq \epsilon + \zeta_i^* \\
& \zeta_i, \zeta_i^*, \epsilon \geq 0
\end{aligned}
$$

Problema Dual

$$
\begin{aligned}
\text{máx} \quad & -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)\left\langle \mathbf{x}_i, \mathbf{x}_j \right\rangle + \sum_{i=1}^{n} y_i(\alpha_i^* + \alpha_i) \\
\text{sujeto a} \quad & \sum_{i=1}^{n}(\alpha_i^* - \alpha_i) = 0 \\
& \alpha_i, \alpha_i^* \in [0, C/n] \\
& \sum_{i=1}^{n}(\alpha_i + \alpha_i^*) \leq C\nu
\end{aligned}
$$

Problema Dual

$$
\begin{aligned}
\text{máx} \quad & -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)\,{\color{red}\langle \mathbf{x}_i, \mathbf{x}_j\rangle} + \sum_{i=1}^{n} y_i(\alpha_i^* + \alpha_i) \\
\text{sujeto a} \quad & \sum_{i=1}^{n}(\alpha_i^* - \alpha_i) = 0 \\
& \alpha_i, \alpha_i^* \in [0, C/n] \\
& \sum_{i=1}^{n}(\alpha_i + \alpha_i^*) \leq C\nu
\end{aligned}
$$

Clasificación de una clase (Detección de anomalías)

Clasificación de una clase (Detección de anomalías)

- Problema: Determinar si un dato x es típico de una población.

Clasificación de una clase (Detección de anomalías)

- Problema: Determinar si un dato x es típico de una población.
- Detectar el conjunto de soporte de una distribución $\mathcal{D}$ tal que $x \sim \mathcal{D}$.

Clasificación de una clase (Detección de anomalías)

- Problema: Determinar si un dato x es típico de una población.
- Detectar el conjunto de soporte de una distribución $\mathcal{D}$ tal que $x \sim \mathcal{D}$.
- Dados datos de entrenamiento $\{x_i\}_{i=1}^n$, queremos encontrar un subconjunto pequeño del espacio de entrada que contenga la mayoría de los datos.

Clasificación de una clase (Detección de anomalías)

- Problema: Determinar si un dato x es típico de una población.
- Detectar el conjunto de soporte de una distribución $\mathcal{D}$ tal que $x \sim \mathcal{D}$.
- Dados datos de entrenamiento $\{x_i\}_{i=1}^n$, queremos encontrar un subconjunto pequeño del espacio de entrada que contenga la mayoría de los datos.
- Clasificador: 1, adentro de la región, -1 fuera de la región.

Clasificación de una clase (Detección de anomalías)

- Problema: Determinar si un dato x es típico de una población.
- Detectar el conjunto de soporte de una distribución $\mathcal{D}$ tal que $x \sim \mathcal{D}$.
- Dados datos de entrenamiento $\{x_i\}_{i=1}^{n}$, queremos encontrar un subconjunto pequeño del espacio de entrada que contenga la mayoría de los datos.
- Clasificador: 1, adentro de la región, -1 fuera de la región.
- Separar datos del orígen con un clasificador lineal.

Clasificación de una clase (Detección de anomalías)

- Problema: Determinar si un dato x es típico de una población.
- Detectar el conjunto de soporte de una distribución $\mathcal{D}$ tal que $x \sim \mathcal{D}$.
- Dados datos de entrenamiento $\{x_i\}_{i=1}^n$, queremos encontrar un subconjunto pequeño del espacio de entrada que contenga la mayoría de los datos.
- Clasificador: 1, adentro de la región, -1 fuera de la región.
- Separar datos del orígen con un clasificador lineal.
- Kernel: proyección no lineal.

- Problema primal:

$$
\begin{aligned}
\min_{\mathbf{w},\zeta_i,{\color{red}\rho}} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{n\nu}\sum_{i=1}^{n}\zeta_i - \rho \\
\text{sujeto a} \quad & \langle \mathbf{w}, \mathbf{x}_i \rangle \geq \rho - \zeta_i \quad i = 1, \ldots, n \\
& \zeta_i \geq 0
\end{aligned}
$$

- Problema primal:

$$\begin{aligned} \min_{\mathbf{w},\zeta_i,\color{red}{\rho}} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{n\nu}\sum_{i=1}^{n}\zeta_i - \rho \\ \text{sujeto a} \quad & \langle \mathbf{w}, \mathbf{x}_i \rangle \geq \rho - \zeta_i \quad i = 1, \ldots, n \\ & \zeta_i \geq 0 \end{aligned}$$

- Problema dual:

$$\begin{aligned} \min \quad & \frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n}\alpha_i\alpha_j \color{red}{\langle x_i, x_j \rangle} \\ \text{sujeto a} \quad & 0 \leq \alpha_i \leq \frac{1}{\nu n} \\ & \sum_{i=1}^{n}\alpha_i = 1 \end{aligned}$$

- Función de decisión (kernelizada):

$$f(x) = \operatorname{sign}\left(\sum_i \alpha_i k(x_i, x) - \rho\right)$$

- Función de decisión (kernelizada):

$$f(x) = \text{sign}\left(\sum_i \alpha_i k(x_i, x) - \rho\right)$$

- Si $\nu \to 0$, α_i no están acotados, y $\rho \Rightarrow -\infty$.

Kernel PCA

Kernel PCA

- PCA:

$$\mathbf{x} = \sum_{i=1}^{L} \langle \mathbf{x}, \mathbf{v}_i \rangle \mathbf{v}_i$$

Kernel PCA

- PCA:

$$\mathbf{x} = \sum_{i=1}^{L} \langle \mathbf{x}, \mathbf{v}_i \rangle \mathbf{v}_i$$

donde $\mathbf{C}\mathbf{v}_i = \lambda_i \mathbf{v}_i$, y

$$\mathbf{C} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j \mathbf{x}_j^T, \quad \sum_{j=1}^{n} \mathbf{x}_j = 0, \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

Kernel PCA

- PCA:

$$\mathbf{x} = \sum_{i=1}^{L} \langle \mathbf{x}, \mathbf{v}_i \rangle \, \mathbf{v}_i$$

donde $\mathbf{C}\mathbf{v}_i = \lambda_i \mathbf{v}_i$, y

$$\mathbf{C} = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j \mathbf{x}_j^T, \quad \sum_{j=1}^{n} \mathbf{x}_j = 0, \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

- Note que:

$$\lambda_i \mathbf{v}_i = \mathbf{C}\mathbf{v}_i$$

Kernel PCA

- PCA:

$$\mathbf{x} = \sum_{i=1}^{L} \langle \mathbf{x}, \mathbf{v}_i \rangle \, \mathbf{v}_i$$

donde $\mathbf{C}\mathbf{v}_i = \lambda_i \mathbf{v}_i$, y

$$\mathbf{C} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j \mathbf{x}_j^T, \quad \sum_{j=1}^{n} \mathbf{x}_j = 0, \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

- Note que:

$$\lambda_i \mathbf{v}_i = \mathbf{C}\mathbf{v}_i = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j \mathbf{x}_j^T \mathbf{v}_i$$

Kernel PCA

- PCA:

$$\mathbf{x} = \sum_{i=1}^{L} \langle \mathbf{x}, \mathbf{v}_i \rangle \mathbf{v}_i$$

donde $\mathbf{C}\mathbf{v}_i = \lambda_i \mathbf{v}_i$, y

$$\mathbf{C} = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j \mathbf{x}_j^T, \quad \sum_{j=1}^{n} \mathbf{x}_j = 0, \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

- Note que:

$$\lambda_i \mathbf{v}_i = \mathbf{C}\mathbf{v}_i = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j \mathbf{x}_j^T \mathbf{v}_i = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j \langle \mathbf{x}_j, \mathbf{v}_i \rangle$$

Kernel PCA

- PCA:

$$\mathbf{x} = \sum_{i=1}^{L} \langle \mathbf{x}, \mathbf{v}_i \rangle \mathbf{v}_i$$

donde $\mathbf{C}\mathbf{v}_i = \lambda_i \mathbf{v}_i$, y

$$\mathbf{C} = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j \mathbf{x}_j^T, \quad \sum_{j=1}^{n} \mathbf{x}_j = 0, \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

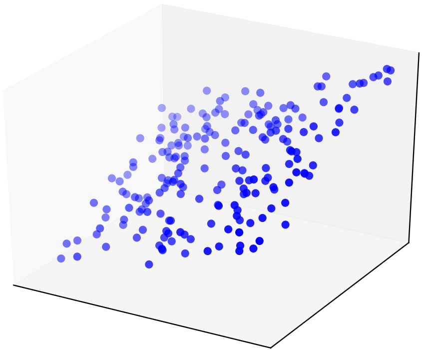
- Note que:

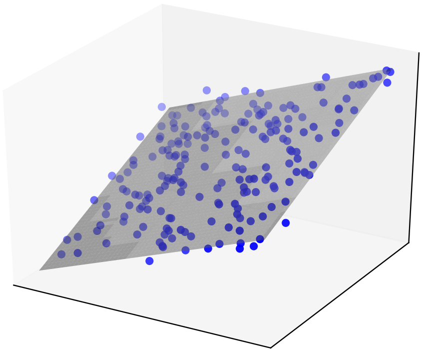
$$\lambda_i \mathbf{v}_i = \mathbf{C}\mathbf{v}_i = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j \mathbf{x}_j^T \mathbf{v}_i = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j \langle \mathbf{x}_j, \mathbf{v}_i \rangle$$

Es decir cada $\mathbf{v}_i \in \operatorname{span}\{\mathbf{x}_i, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$

- Sistema de ecuaciones:

$$\lambda \langle \mathbf{x}_j, \mathbf{v}_i \rangle = \langle \mathbf{x}_j, \mathbf{C}\mathbf{v}_i \rangle \quad j = 1, \ldots, n$$

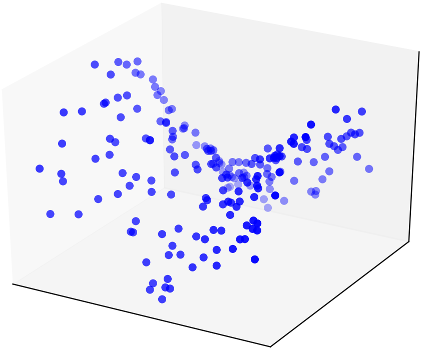




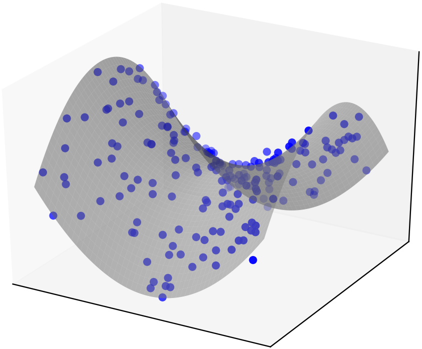
Estructura de los datos está en subespacio de menor dimensión

- Estructura puede ser no lineal:

- Estructura puede ser no lineal:



- Estructura puede ser no lineal:



- Kernel PCA es PCA en espacio de características:

- Kernel PCA es PCA en espacio de características:

$$\begin{aligned}\mathcal{X} &\to \mathcal{H}\\ x &\mapsto \phi(x)\end{aligned}$$

- Kernel PCA es PCA en espacio de características:

$$\mathcal{X} \to \mathcal{H}$$
$$x \mapsto \phi(x)$$

- Suponemos datos centrados:

$$\sum_{i=1}^{n} \phi(x_i) = 0$$

- Kernel PCA es PCA en espacio de características:

$$\mathcal{X} \to \mathcal{H}$$
$$x \mapsto \boldsymbol{\phi}(x)$$

- Suponemos datos centrados:

$$\sum_{i=1}^{n} \boldsymbol{\phi}(x_i) = 0$$

- Consideramos matriz de covarianza:

$$\mathbf{C} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\phi}(x_i) \boldsymbol{\phi}(x_i)^T$$

- Kernel PCA es PCA en espacio de características:

$$\mathcal{X} \to \mathcal{H}$$
$$x \mapsto \boldsymbol{\phi}(x)$$

- Suponemos datos centrados:

$$\sum_{i=1}^{n} \boldsymbol{\phi}(x_i) = 0$$

- Consideramos matriz de covarianza:

$$\mathbf{C} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\phi}(x_i) \boldsymbol{\phi}(x_i)^T$$

- Si $\dim(\mathcal{H}) = \infty$, $\boldsymbol{\phi}(x_i)\boldsymbol{\phi}(x_i)^T$ corresponde a operador lineal en $\mathcal{H}$:

$$\mathbf{x} \mapsto \boldsymbol{\phi}(x_i) \langle \boldsymbol{\phi}(x_i), \mathbf{x} \rangle_{\mathcal{H}}$$

- Queremos encontrar valores y vectores propios:

$$\lambda \mathbf{v} = \mathbf{C}\mathbf{v}$$

- Queremos encontrar valores y vectores propios:

$$\lambda \mathbf{v} = \mathbf{C}\mathbf{v}$$

- Ecuaciones:

- Queremos encontrar valores y vectores propios:

$$\lambda \mathbf{v} = \mathbf{C}\mathbf{v}$$

- Ecuaciones:

$$\lambda \left\langle \boldsymbol{\phi}(x_j), \mathbf{v} \right\rangle_{\mathcal{H}} = \left\langle \boldsymbol{\phi}(x_j), \mathbf{C}\mathbf{v} \right\rangle_{\mathcal{H}} \quad j = 1, \ldots, n$$

- Queremos encontrar valores y vectores propios:

$$\lambda \mathbf{v} = \mathbf{C}\mathbf{v}$$

- Ecuaciones:

$$\lambda \left\langle \boldsymbol{\phi}(x_j), \mathbf{v} \right\rangle_{\mathcal{H}} = \left\langle \boldsymbol{\phi}(x_j), \mathbf{C}\mathbf{v} \right\rangle_{\mathcal{H}} \quad j = 1, \ldots, n$$

$$\mathbf{v} = \sum_{i=1}^{n} \alpha_i \boldsymbol{\phi}(x_i)$$

- Queremos encontrar valores y vectores propios:

$$\lambda \mathbf{v} = \mathbf{C}\mathbf{v}$$

- Ecuaciones:

$$\lambda \langle \boldsymbol{\phi}(x_j), \mathbf{v} \rangle_{\mathcal{H}} = \langle \boldsymbol{\phi}(x_j), \mathbf{C}\mathbf{v} \rangle_{\mathcal{H}} \quad j = 1, \ldots, n$$

$$\mathbf{v} = \sum_{i=1}^{n} \alpha_i \boldsymbol{\phi}(x_i)$$

- Combinando:

$$\lambda \sum_{i=1}^{n} \alpha_i \langle \boldsymbol{\phi}(x_j), \boldsymbol{\phi}(x_i) \rangle_{\mathcal{H}} =$$

$$\frac{1}{n} \sum_{i=1}^{n} \alpha_i \left\langle \boldsymbol{\phi}(x_j), \sum_{j=1}^{n} \boldsymbol{\phi}(x_j) \langle \boldsymbol{\phi}(x_j), \boldsymbol{\phi}(x_i) \rangle_{\mathcal{H}} \right\rangle_{\mathcal{H}}$$

para $j = 1, \ldots, n$

- En términos de la matriz de Gramm, es equivalente a:

$$n\lambda \mathbf{K}\boldsymbol{\alpha} = \mathbf{K}^2\boldsymbol{\alpha}$$

- En términos de la matriz de Gramm, es equivalente a:

$$n\lambda \mathbf{K}\boldsymbol{\alpha} = \mathbf{K}^2\boldsymbol{\alpha}$$

cuyas soluciones se pueden encontrar resolviendo:

$$n\lambda \boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha}$$

- En términos de la matriz de Gramm, es equivalente a:

$$n\lambda \mathbf{K}\boldsymbol{\alpha} = \mathbf{K}^2\boldsymbol{\alpha}$$

cuyas soluciones se pueden encontrar resolviendo:

$$n\lambda \boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha}$$

- Normalización $\|\mathbf{v}\|_{\mathcal{H}} = 1$:

- En términos de la matriz de Gramm, es equivalente a:

$$n\lambda \mathbf{K}\boldsymbol{\alpha} = \mathbf{K}^2\boldsymbol{\alpha}$$

cuyas soluciones se pueden encontrar resolviendo:

$$n\lambda\boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha}$$

- Normalización $\|\mathbf{v}\|_{\mathcal{H}} = 1$:

$$1 = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i, \alpha_j \left\langle \boldsymbol{\phi}(x_j), \boldsymbol{\phi}(x_i) \right\rangle$$

- En términos de la matriz de Gramm, es equivalente a:

$$n\lambda\mathbf{K}\boldsymbol{\alpha} = \mathbf{K}^2\boldsymbol{\alpha}$$

cuyas soluciones se pueden encontrar resolviendo:

$$n\lambda\boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha}$$

- Normalización $\|\mathbf{v}\|_{\mathcal{H}} = 1$:

$$1 = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i, \alpha_j \left\langle \boldsymbol{\phi}(x_j), \boldsymbol{\phi}(x_i)\right\rangle = \langle\boldsymbol{\alpha}, \mathbf{K}\boldsymbol{\alpha}\rangle$$

- En términos de la matriz de Gramm, es equivalente a:

$$n\lambda \mathbf{K}\boldsymbol{\alpha} = \mathbf{K}^2\boldsymbol{\alpha}$$

cuyas soluciones se pueden encontrar resolviendo:

$$n\lambda\boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha}$$

- Normalización $\|\mathbf{v}\|_{\mathcal{H}} = 1$:

$$1 = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i, \alpha_j \left\langle \boldsymbol{\phi}(x_j), \boldsymbol{\phi}(x_i) \right\rangle = \left\langle \boldsymbol{\alpha}, \mathbf{K}\boldsymbol{\alpha} \right\rangle = \lambda \left\langle \boldsymbol{\alpha}, \boldsymbol{\alpha} \right\rangle$$

- En términos de la matriz de Gramm, es equivalente a:

$$n\lambda \mathbf{K}\boldsymbol{\alpha} = \mathbf{K}^2\boldsymbol{\alpha}$$

cuyas soluciones se pueden encontrar resolviendo:

$$n\lambda\boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha}$$

- Normalización $\|\mathbf{v}\|_{\mathcal{H}} = 1$:

$$1 = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i, \alpha_j \left\langle \boldsymbol{\phi}(x_j), \boldsymbol{\phi}(x_i) \right\rangle = \left\langle \boldsymbol{\alpha}, \mathbf{K}\boldsymbol{\alpha} \right\rangle = \lambda \left\langle \boldsymbol{\alpha}, \boldsymbol{\alpha} \right\rangle$$

- Para un nuevo dato x:

- En términos de la matriz de Gramm, es equivalente a:

$$n\lambda \mathbf{K}\boldsymbol{\alpha} = \mathbf{K}^2\boldsymbol{\alpha}$$

cuyas soluciones se pueden encontrar resolviendo:

$$n\lambda \boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha}$$

- Normalización $\|\mathbf{v}\|_{\mathcal{H}} = 1$:

$$1 = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i, \alpha_j \left\langle \boldsymbol{\phi}(x_j), \boldsymbol{\phi}(x_i) \right\rangle = \left\langle \boldsymbol{\alpha}, \mathbf{K}\boldsymbol{\alpha} \right\rangle = \lambda \left\langle \boldsymbol{\alpha}, \boldsymbol{\alpha} \right\rangle$$

- Para un nuevo dato x:

$$\left\langle \mathbf{v}, \boldsymbol{\phi}(x) \right\rangle_{\mathcal{H}}$$

- En términos de la matriz de Gramm, es equivalente a:

$$n\lambda \mathbf{K}\boldsymbol{\alpha} = \mathbf{K}^2\boldsymbol{\alpha}$$

  cuyas soluciones se pueden encontrar resolviendo:

$$n\lambda\boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha}$$

- Normalización $\|\mathbf{v}\|_{\mathcal{H}} = 1$:

$$1 = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i, \alpha_j \left\langle \boldsymbol{\phi}(x_j), \boldsymbol{\phi}(x_i) \right\rangle = \left\langle \boldsymbol{\alpha}, \mathbf{K}\boldsymbol{\alpha} \right\rangle = \lambda \left\langle \boldsymbol{\alpha}, \boldsymbol{\alpha} \right\rangle$$

- Para un nuevo dato x:

$$\left\langle \mathbf{v}, \boldsymbol{\phi}(x) \right\rangle_{\mathcal{H}} = \sum_{i=1}^{n} \alpha_i \left\langle \boldsymbol{\phi}(x_i), \boldsymbol{\phi}(x) \right\rangle_{\mathcal{H}}$$

- En términos de la matriz de Gramm, es equivalente a:

$$n\lambda \mathbf{K}\boldsymbol{\alpha} = \mathbf{K}^2\boldsymbol{\alpha}$$

cuyas soluciones se pueden encontrar resolviendo:

$$n\lambda\boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha}$$

- Normalización $\|\mathbf{v}\|_{\mathcal{H}} = 1$:

$$1 = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i, \alpha_j \langle \boldsymbol{\phi}(x_j), \boldsymbol{\phi}(x_i) \rangle = \langle \boldsymbol{\alpha}, \mathbf{K}\boldsymbol{\alpha} \rangle = \lambda \langle \boldsymbol{\alpha}, \boldsymbol{\alpha} \rangle$$

- Para un nuevo dato x:

$$\langle \mathbf{v}, \boldsymbol{\phi}(x) \rangle_{\mathcal{H}} = \sum_{i=1}^{n} \alpha_i \langle \boldsymbol{\phi}(x_i), \boldsymbol{\phi}(x) \rangle_{\mathcal{H}} = \sum_{i=1}^{n} \alpha_i k(x_i, x)$$

- Para obtener PCA con datos centrados en $\mathcal{H}$, reemplazanos $\mathbf{K}$ por:

$$\tilde{\mathbf{K}} = \mathbf{K} - \mathbf{JK} - \mathbf{KJ} + \mathbf{JKJ}$$

  donde $[\mathbf{J}]_{ij} = \frac{1}{n}, \forall i, j$