Máquina de Soporte Vectorial en paralelo.

Clasificación.

Clasificación.

- Modelo lineal de clasificación.

Clasificación.

- Modelo lineal de clasificación.
- Sean $\mathcal{C}_0, \mathcal{C}_1$ dos clases y $\mathbf{x} \in \mathbb{R}^n$ un vector de atributos. Se desea clasificar al vector $\mathbf{x}$ en alguna de las dos clases ajenas de modo que se tenga el menor error de clasificación.

Clasificación.

- Modelo lineal de clasificación.
- Sean $\mathcal{C}_0, \mathcal{C}_1$ dos clases y $\mathbf{x} \in \mathbb{R}^n$ un vector de atributos. Se desea clasificar al vector $\mathbf{x}$ en alguna de las dos clases ajenas de modo que se tenga el menor error de clasificación.
- Problemas de este tipo surgen naturalmente en medicina (cáncer), finanzas (crédito), computación (spam),...

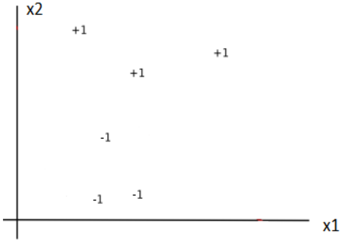
MSV.

MSV.

- Supóngase que se tienen m individuos a los que se desea clasificar en $\mathcal{C}_0, \mathcal{C}_1$ y para cada individuo sean $\mathbf{x}_i \in \mathbb{R}^n$ su correspondiente vector de atributos $i = 1, 2, \ldots, m$ y el vector de etiquetas $\hat{\mathbf{y}} = (\hat{y}_i)_{i=1}^m \in \mathbb{R}^m$, $\hat{y}_i \in \{-1, 1\}$.

MSV.

- Supóngase que se tienen m individuos a los que se desea clasificar en $\mathcal{C}_0, \mathcal{C}_1$ y para cada individuo sean $\mathbf{x}_i \in \mathbb{R}^n$ su correspondiente vector de atributos $i = 1, 2, \ldots, m$ y el vector de etiquetas $\hat{\mathbf{y}} = (\hat{y}_i)_{i=1}^m \in \mathbb{R}^m$, $\hat{y}_i \in \{-1, 1\}$.
- En $\mathbb{R}^2$:



MSV.

MSV.

- Modelo para MSV:

MSV.

- Modelo para MSV:

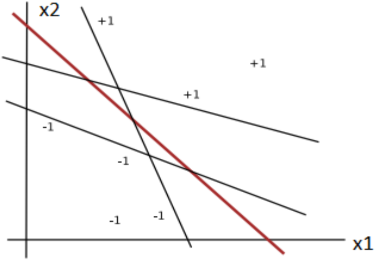
$$y(\mathbf{x}|w_0, \mathbf{w}) = \mathbf{w}^T\mathbf{x} + w_0$$

MSV.

- Modelo para MSV:

$$y(\mathbf{x}|w_0, \mathbf{w}) = \mathbf{w}^T \mathbf{x} + w_0$$

- En $\mathbb{R}^2$:



MSV en $\mathbb{R}^2$.

MSV en $\mathbb{R}^2$.

- En MSV se construye la recta $\mathcal{P} = \{\mathbf{x} \in \mathbb{R}^2 \quad | \quad y(\mathbf{x}|w_0, \mathbf{w}) = 0\}$ que cumpla con dos condiciones:

MSV en $\mathbb{R}^2$.

- En MSV se construye la recta $\mathcal{P} = \{\mathbf{x} \in \mathbb{R}^2 \quad | \quad y(\mathbf{x}|w_0, \mathbf{w}) = 0\}$ que cumpla con dos condiciones:
  - Tenga distancia mínima $d(\cdot, \cdot)$ al conjunto de individuos:

$$d(\mathcal{P}, \mathbf{x}_i) = \frac{|y(\mathbf{x}_i|w_0, \mathbf{w})|}{||\mathbf{w}||} = \frac{\hat{y}_i y_i}{||\mathbf{w}||}$$

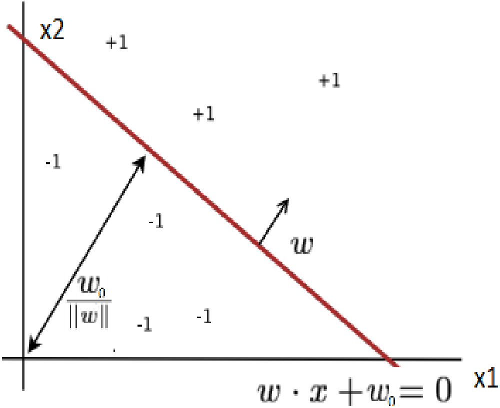
MSV en $\mathbb{R}^2$.

- En MSV se construye la recta $\mathcal{P} = \{\mathbf{x} \in \mathbb{R}^2 \quad | \quad y(\mathbf{x}|w_0, \mathbf{w}) = 0\}$ que cumpla con dos condiciones:
  - Tenga distancia mínima $d(\cdot, \cdot)$ al conjunto de individuos:
    $$d(\mathcal{P}, \mathbf{x}_i) = \frac{|y(\mathbf{x}_i|w_0, \mathbf{w})|}{||\mathbf{w}||} = \frac{\hat{y}_i y_i}{||\mathbf{w}||}$$
  - Su margen sea máximo, considerando el margen $\mathcal{M}_{w_0,\mathbf{w}}$ como:
    $$\mathcal{M}_{w_0,\mathbf{w}} = \min_{i \in \{1,\ldots,m\}} \{d(\mathcal{P}, \mathbf{x}_i)\}$$

MSV en $\mathbb{R}^2$.

MSV en $\mathbb{R}^2$.

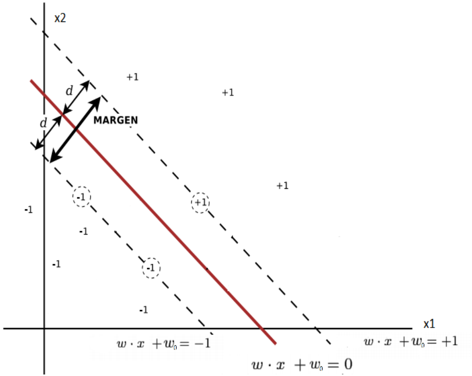
- Se obtiene para este ejemplo la recta siguiente:



MSV en $\mathbb{R}^2$.

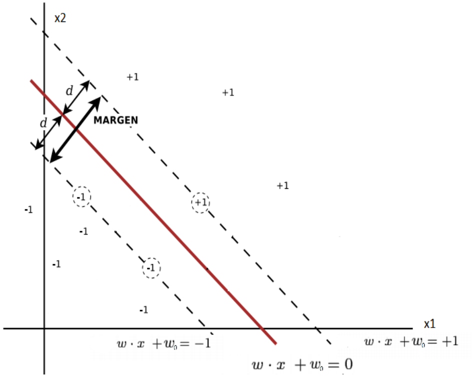
MSV en $\mathbb{R}^2$.

- Se obtiene para este ejemplo la recta siguiente:



MSV en $\mathbb{R}^2$.

- Se obtiene para este ejemplo la recta siguiente:



- Los individuos encerrados en círculos representan los vectores de soporte de la recta.

MSV.

MSV.

- Un individuo $(\mathbf{x}_i, \hat{y}_i)$ está bien clasificado si se verifica:
  - $\mathbf{w}^T\mathbf{x}_i + w_0 \geq 1$ para $\hat{y}_i = 1$
  - $\mathbf{w}^T\mathbf{x}_i + w_0 \leq -1$ para $\hat{y}_i = -1$

MSV.

- Un individuo $(\mathbf{x}_i, \hat{y}_i)$ está bien clasificado si se verifica:
  - $\mathbf{w}^T\mathbf{x}_i + w_0 \geq 1$ para $\hat{y}_i = 1$
  - $\mathbf{w}^T\mathbf{x}_i + w_0 \leq -1$ para $\hat{y}_i = -1$
- Ambas condiciones se escriben:

$$\hat{y}_i(w^T\mathbf{x}_i + w_0) \geq 1 \quad \forall i = 1, 2, \ldots, m$$

MSV.

MSV.

- Suponiendo que los datos son linealmente separables en el espacio $\mathbb{R}^n$ el problema que se debe resolver es:

MSV.

- Suponiendo que los datos son linealmente separables en el espacio $\mathbb{R}^n$ el problema que se debe resolver es:

$$\min_{w_0,\mathbf{w}} \frac{1}{2}||(w_0,\mathbf{w})||^2 \quad \text{sujeto a:} \quad \hat{y}_i y_i \geq 1 \quad \forall i = 1,\ldots,m$$

MSV.

- Suponiendo que los datos son linealmente separables en el espacio $\mathbb{R}^n$ el problema que se debe resolver es:

$$\min_{w_0, \mathbf{w}} \frac{1}{2} ||(w_0, \mathbf{w})||^2 \quad \text{sujeto a:} \quad \hat{y}_i y_i \geq 1 \quad \forall i = 1, \ldots, m$$

- La regla para un nuevo individuo $\mathbf{x}_N$ es: si $y(\mathbf{x}_N | \mathbf{w}, w_0) \geq 0$ entonces se clasifica a $\mathcal{C}_0$ y en otro caso a $\mathcal{C}_1$

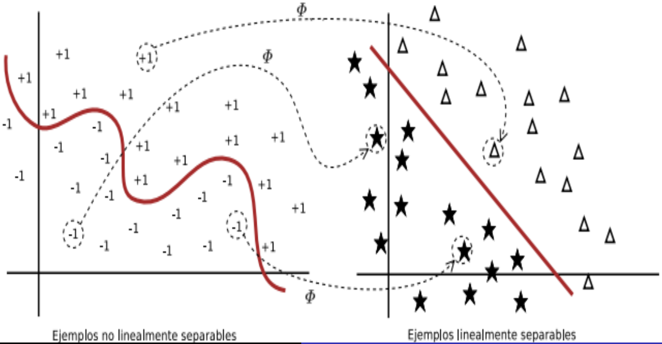
Problema dual.

Problema dual.

- La formulación dual del problema de optimización anterior tiene ventajas en la práctica por la posibilidad de usar funciones que permitan separar linealmente a los individuos.

Problema dual.

- La formulación dual del problema de optimización anterior tiene ventajas en la práctica por la posibilidad de usar funciones que permitan separar linealmente a los individuos.



Problema dual.

Problema dual.

- El problema dual es:

$$\min_{\lambda} \frac{1}{2}\lambda^T \hat{Y} X^T X \hat{Y} \lambda - \lambda^T e \quad \text{sujeto a:} \quad e^T \hat{Y} \lambda = 0, \quad \lambda \geq 0$$

donde:

Problema dual.

- El problema dual es:

$$\min_{\lambda} \frac{1}{2}\lambda^T \hat{Y} X^T X \hat{Y} \lambda - \lambda^T e \quad \text{sujeto a:} \quad e^T \hat{Y} \lambda = 0, \quad \lambda \geq 0$$

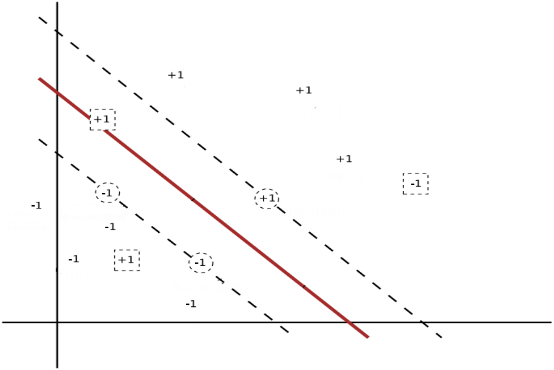
donde:

  - $e = (1)_{i=1}^{m}$
  - $X = \left[\begin{array}{ccc} \mathbf{x}_1 & \cdots & \mathbf{x}_m \end{array} \right]$
  - $\hat{Y} = \operatorname{diag}(\hat{y}_i)_{i=1}^{m}$
  - $\lambda \in \mathbb{R}^m$ vector de multiplicadores de Lagrange.

MSV en $\mathbb{R}^2$

MSV en $\mathbb{R}^2$

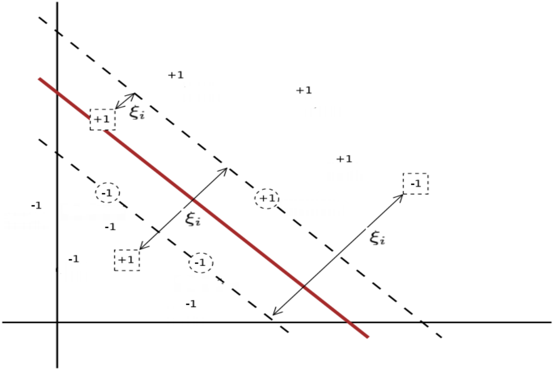
- En el caso que se tenga una situación del tipo:



MSV en $\mathbb{R}^2$

MSV en $\mathbb{R}^2$

- Se definen para cada individuo $\xi_i \geq 0 \quad \forall i = 1, 2, \ldots, m$:



MSV.

MSV.

- El problema de optimización en este caso:

$$\min_{w_0,\mathbf{w}} \frac{1}{2}||(w_0, \mathbf{w})||^2 + Ce^T\boldsymbol{\xi}$$

$$\text{sujeto a:} \quad \boldsymbol{\xi} \geq 0 \qquad \hat{y}_i y_i \geq 1 - \xi_i \quad \forall i = 1, \ldots, m$$

MSV.

- El problema de optimización en este caso:

$$\min_{w_0, \mathbf{w}} \frac{1}{2} ||(w_0, \mathbf{w})||^2 + Ce^T \boldsymbol{\xi}$$

$$\text{sujeto a:} \quad \boldsymbol{\xi} \geq 0 \qquad \hat{y}_i y_i \geq 1 - \xi_i \quad \forall i = 1, \ldots, m$$

donde: $C > 0$ realiza un equilibrio entre sobreajuste y complejidad del modelo.

MSV dual.

MSV dual.

- El problema de optimización dual (POD):

$$\min_{\lambda_g} \frac{1}{2} \lambda_g^T \hat{Y} X^T X \hat{Y} \lambda_g - \lambda_g^T e$$

$$\text{sujeto a:} \quad e^T \hat{Y} \lambda_g = 0, \quad 0 \leq \lambda_g \leq C$$

MSV dual.

- El problema de optimización dual (POD):

$$\min_{\lambda_g} \frac{1}{2}\lambda_g^T \hat{Y} X^T X \hat{Y} \lambda_g - \lambda_g^T e$$

$$\text{sujeto a:} \quad e^T \hat{Y} \lambda_g = 0, \quad 0 \leq \lambda_g \leq C$$

donde $e, X, \hat{Y}$ son como antes y $\lambda_g \in \mathbb{R}^m$ corresponde al vector de multiplicadores de Lagrange para la restricciones de la forma $\hat{y}_i y_i \geq 1 - \xi_i \quad \forall i = 1, 2, \ldots, m$

Puntos interiores

Puntos interiores

- Ambas formulaciones duales son problemas de optimización cuya función es cuadrática con restricciones lineales.

Puntos interiores

- Ambas formulaciones duales son problemas de optimización cuya función es cuadrática con restricciones lineales.
- Métodos convencionales para resolver estos problemas han sido propuestos por el análisis numérico.

Puntos interiores

- Ambas formulaciones duales son problemas de optimización cuya función es cuadrática con restricciones lineales.
- Métodos convencionales para resolver estos problemas han sido propuestos por el análisis numérico.
- Entre ellos se encuentran los métodos por puntos interiores.

Puntos interiores

- Ambas formulaciones duales son problemas de optimización cuya función es cuadrática con restricciones lineales.
- Métodos convencionales para resolver estos problemas han sido propuestos por el análisis numérico.
- Entre ellos se encuentran los métodos por puntos interiores.
- La formulación por puntos interiores que veremos se realizará con el enfoque de función de barrera logarítmica.

POD

POD

- Realizando una asignación de variables y agregando la holgura s para la condición $\lambda_g \leq Ce$ se tiene el problema de optimización:

POD

- Realizando una asignación de variables y agregando la holgura s para la condición $\lambda_g \leq Ce$ se tiene el problema de optimización:

$$\min_{x,y,s} \quad \frac{y^T y}{2} - e^T x \quad \text{sujeto a:}$$

POD

- Realizando una asignación de variables y agregando la holgura s para la condición $\lambda_g \leq Ce$ se tiene el problema de optimización:

$$\min_{x,y,s} \quad \frac{y^T y}{2} - e^T x \quad \text{sujeto a:}$$

$$\begin{aligned} b^T x &= 0 \\ Ax - y &= 0 \\ x - Ce + s &= 0 \\ x &\geq 0 \\ s &\geq 0. \end{aligned}$$

POD con función de barrera

POD con función de barrera

- Utilizando la función de barrera logarítmica:

POD con función de barrera

- Utilizando la función de barrera logarítmica:

$$\min_{x,y,s} \quad \frac{y^T y}{2} - e^T x - \mu_x \sum_{i=1}^{m} \log x_i - \mu_s \sum_{i=1}^{m} \log s_i \quad \text{sujeto a:}$$

$$\begin{aligned} b^T x &= 0 \\ Ax - y &= 0 \\ x - Ce + s &= 0. \end{aligned}$$

Condiciones KKT para MSV secuencial

Condiciones KKT para MSV secuencial

- Las condiciones KKT del problema anterior conducen a un sistema lineal de la forma:

$$Kd = -r(\mu),$$

Condiciones KKT para MSV secuencial

- Las condiciones KKT del problema anterior conducen a un sistema lineal de la forma:

$$Kd = -r(\mu),$$

con $K = \begin{bmatrix} X^{-1}Z + S^{-1}W & A^T & 0 & -b \\ A & -I_n & 0 & 0 \\ I_m & 0 & I_m & 0 \\ b^T & 0 & 0 & 0 \end{bmatrix}$, r conocidos, $\mu = \begin{bmatrix} \mu_x \\ \mu_s \end{bmatrix}$.

POD con función de barrera

POD con función de barrera

- Algunas consideraciones para el problema anterior:

POD con función de barrera

- Algunas consideraciones para el problema anterior:
  - La solución del problema se obtiene haciendo $\mu_x, \mu_s \to 0$ en cada iteración.

POD con función de barrera

- Algunas consideraciones para el problema anterior:
    - La solución del problema se obtiene haciendo $\mu_x, \mu_s \to 0$ en cada iteración.
    - Esta solución es única.

POD con función de barrera

- Algunas consideraciones para el problema anterior:
  - La solución del problema se obtiene haciendo $\mu_x, \mu_s \to 0$ en cada iteración.
  - Esta solución es única.
  - Se agrega la corrección de Mehrotra para acelerar convergencia.

POD con función de barrera

- Algunas consideraciones para el problema anterior:
  - La solución del problema se obtiene haciendo $\mu_x, \mu_s \to 0$ en cada iteración.
  - Esta solución es única.
  - Se agrega la corrección de Mehrotra para acelerar convergencia.
  - La matriz K no posee una estructura para una formulación en paralelo....

POD con función de barrera

- Algunas consideraciones para el problema anterior:
  - La solución del problema se obtiene haciendo $\mu_x, \mu_s \to 0$ en cada iteración.
  - Esta solución es única.
  - Se agrega la corrección de Mehrotra para acelerar convergencia.
  - La matriz K no posee una estructura para una formulación en paralelo.... solución: *Using interior point methods for large scale support vector machine training* por K. Woodsend y J. Gondzio, 2009.

POD (2) con función de barrera

POD (2) con función de barrera

- Se realiza la misma asignación de variables pero no se agrega la holgura s:

POD (2) con función de barrera

- Se realiza la misma asignación de variables pero no se agrega la holgura s:

$$\min_{x,y} \quad \frac{y^T y}{2} - e^T x - \mu \left[\sum_{i=1}^{m} \log x_i + \log \left(Ce_i - x_i \right) \right] \quad \text{sujeto a:}$$

$$A_y y + A_x x = 0$$

POD (2) con función de barrera

- Se realiza la misma asignación de variables pero no se agrega la holgura s:

$$\min_{x,y} \quad \frac{y^T y}{2} - e^T x - \mu \left[\sum_{i=1}^{m} \log x_i + \log \left(Ce_i - x_i \right) \right] \quad \text{sujeto a:}$$

$$A_y y + A_x x = 0$$

donde: $A_y = \begin{bmatrix} I_n \\ 0 \end{bmatrix}$, $\quad A_x = \begin{bmatrix} -A \\ -b^T \end{bmatrix}$, $\mu > 0$ es el parámetro de barrera logarítmica.

Condiciones KKT para MSV en paralelo

Condiciones KKT para MSV en paralelo

- Las condiciones de KKT del problema anterior conducen a un sistema lineal de la forma:

$$Kd = -r$$

Condiciones KKT para MSV en paralelo

- Las condiciones de KKT del problema anterior conducen a un sistema lineal de la forma:

$$Kd = -r$$

donde: K es una matriz simétrica, indefinida cuyas entradas son conocidas y $r = \begin{bmatrix} rc_y \\ rc_x(\mu) \\ r_b \end{bmatrix}$

Condiciones KKT para MSV en paralelo

- Las condiciones de KKT del problema anterior conducen a un sistema lineal de la forma:

$$Kd = -r$$

  donde: K es una matriz simétrica, indefinida cuyas entradas son conocidas y $r = \begin{bmatrix} rc_y \\ rc_x(\mu) \\ r_b \end{bmatrix}$

- El óptimo del problema anterior se encuentra haciendo $\mu \to 0$ en cada iteración.

Puntos interiores en paralelo

Puntos interiores en paralelo

- Una ventaja del sistema anterior tiene que ver con la estructura de la matriz K:

$$K = \left[\begin{array}{cc} H & -\hat{A}^T \\ -\hat{A} & 0 \end{array} \right]$$

Puntos interiores en paralelo

- Una ventaja del sistema anterior tiene que ver con la estructura de la matriz K:

$$K = \begin{bmatrix} H & -\hat{A}^T \\ -\hat{A} & 0 \end{bmatrix}$$

con H una matriz diagonal, $\hat{A}$ rectangular.

Puntos interiores en paralelo

- Una ventaja del sistema anterior tiene que ver con la estructura de la matriz K:
$$K = \begin{bmatrix} H & -\hat{A}^T \\ -\hat{A} & 0 \end{bmatrix}$$
con H una matriz diagonal, $\hat{A}$ rectangular.
- Si se tienen p procesadores, entonces K se puede escribir como sigue: $K = \begin{bmatrix} H_1 & & & & A_1^T \\ & H_2 & & & A_2^T \\ & & \ddots & & \vdots \\ & & & H_p & A_p^T \\ A_1 & A_2 & \cdots & A_p & 0 \end{bmatrix}$

Puntos interiores en paralelo

Puntos interiores en paralelo

- Por otro lado, existe una factorización para K de la forma LD^TL:

Puntos interiores en paralelo

- Por otro lado, existe una factorización para K de la forma LD^TL:

$$L = \begin{bmatrix} L_1 & & & \\ & \ddots & & \\ & & L_p & \\ L_{A_1} & \cdots & L_{A_p} & L_C \end{bmatrix}, \quad D = \begin{bmatrix} D_1 & & & \\ & \ddots & & \\ & & D_p & \\ & & & D_C \end{bmatrix}$$

Puntos interiores en paralelo

Puntos interiores en paralelo

- Se concluye lo siguiente:

$$
\begin{aligned}
H_i &= L_i D_i L_i^T, \quad \therefore D_i = H_i, \quad L_i = I && (1)\\
L_{A_i} &= A_i L_i^{-T} D_i^{-1} = A_i H_i^{-1} && (2)\\
C &\equiv -\sum_{i=1}^{p} A_i H_i^{-1} A_i^T && (3)\\
&= L_C D_C L_C^T. && (4)
\end{aligned}
$$

Puntos interiores en paralelo

- Se concluye lo siguiente:

$$
\begin{aligned}
H_i &= L_i D_i L_i^T, \quad \therefore D_i = H_i, \quad L_i = I && (1)\\
L_{A_i} &= A_i L_i^{-T} D_i^{-1} = A_i H_i^{-1} && (2)\\
C &\equiv -\sum_{i=1}^{p} A_i H_i^{-1} A_i^T && (3)\\
&= L_C D_C L_C^T. && (4)
\end{aligned}
$$

Los productos (2) se realizan en cada uno de los procesadores y la suma (3) y la factorización (4) se realiza en un sólo procesador.

Puntos interiores en paralelo

Puntos interiores en paralelo

- Con la factorización $K = LDL^T$ se resuelve el sistema:

Puntos interiores en paralelo

- Con la factorización $K = LDL^T$ se resuelve el sistema:

$$LDL^T \begin{bmatrix} d_z \\ d_\lambda \end{bmatrix} = - \begin{bmatrix} r_c \\ r_b \end{bmatrix},$$

donde $z = \begin{bmatrix} y \\ x \end{bmatrix}$, $rc = \begin{bmatrix} rc_y \\ rc_x \end{bmatrix}$ también en paralelo.

Puntos interiores en paralelo

Puntos interiores en paralelo

- Se tiene:

$$
\begin{aligned}
d\lambda'' &= L_C^{-1}\left(r_b - \sum_{i=1}^{p} L_{A_i} r_{c_i}\right) && (5)\\
d\lambda' &= D_C^{-1} d\lambda'' && (6)\\
d\lambda &= L_C^{-T} d\lambda' && (7)\\
dz_i' &= D_i^{-1} r_{c_i} && (8)\\
dz_i &= dz_i' - L_{A_i}^{T} d\lambda. && (9)
\end{aligned}
$$

Puntos interiores en paralelo

- Se tiene:

$$
\begin{aligned}
d\lambda'' &= L_C^{-1}\left(r_b - \sum_{i=1}^{p} L_{A_i} r_{c_i}\right) && (5)\\
d\lambda' &= D_C^{-1} d\lambda'' && (6)\\
d\lambda &= L_C^{-T} d\lambda' && (7)\\
dz_i' &= D_i^{-1} r_{c_i} && (8)\\
dz_i &= dz_i' - L_{A_i}^T d\lambda. && (9)
\end{aligned}
$$

En cada procesador se lleva a cabo el producto $L_{A_i} r_{c_i}$ y en un sólo procesador se realiza (5), (6), (7). El vector $d\lambda$ se transmite a cada procesador en los que se realiza el producto $D_i^{-1} r_{c_i}$ y (9).

Bases de datos

Bases de datos

- Criterios de selección:

Bases de datos

- Criterios de selección:
  - Dimensiones.

Bases de datos

- Criterios de selección:
  - Dimensiones.
  - Sin valores faltantes.

Bases de datos

- Criterios de selección:
  - Dimensiones.
  - Sin valores faltantes.
  - De acuerdo a las características de la computadora (procesador con dos núcleos y 8 gb en RAM)

Bases de datos

- Criterios de selección:
  - Dimensiones.
  - Sin valores faltantes.
  - De acuerdo a las características de la computadora (procesador con dos núcleos y 8 gb en RAM)
  - Se elige la tasa de clasificación incorrecta como medida de precisión.

Avisos de granja

Avisos de granja

- Avisos de texto en 12 sitios de internet.

Avisos de granja

- Avisos de texto en 12 sitios de internet.
- Modelo por bolsa de palabras.

Avisos de granja

- Avisos de texto en 12 sitios de internet.
- Modelo por bolsa de palabras.
- Etiquetas binarias (aprueba o no el aviso)

Avisos de granja

- Avisos de texto en 12 sitios de internet.
- Modelo por bolsa de palabras.
- Etiquetas binarias (aprueba o no el aviso)
- 4,143 renglones y 54,877 columnas.

Avisos de granja

- Avisos de texto en 12 sitios de internet.
- Modelo por bolsa de palabras.
- Etiquetas binarias (aprueba o no el aviso)
- 4,143 renglones y 54,877 columnas.
- Se divide en un 10 % para el conjunto de prueba

Avisos de granja

	MSV	RL
Entrenamiento	1.072e-3	2.681 e-3
Prueba	7.729e-2	5.797e-2

Trombina

Trombina

- Compuestos presentes en la sangre.

Trombina

- Compuestos presentes en la sangre.
- Etiquetas binarias (el compuesto se une o no a la trombina)

Trombina

- Compuestos presentes en la sangre.
- Etiquetas binarias (el compuesto se une o no a la trombina)
- El conjunto de entrenamiento tiene 1,909 renglones y 139,351 columnas

Trombina

- Compuestos presentes en la sangre.
- Etiquetas binarias (el compuesto se une o no a la trombina)
- El conjunto de entrenamiento tiene 1,909 renglones y 139,351 columnas
- El conjunto de prueba tiene 634 renglones.

Trombina

	MSV	RL
Entrenamiento	1.047e-3	2.2e-2
Prueba	2.3659e-1	2.3659e-1

Comentarios.

Comentarios.

- La conexión entre nodos de un clúster, la transmisión de datos entre ellos y el manejo de memoria es posible consultar en la referencia de Woodsend y Gondzio.

Comentarios.

- La conexión entre nodos de un clúster, la transmisión de datos entre ellos y el manejo de memoria es posible consultar en la referencia de Woodsend y Gondzio.
- La formulación dual de la MSV permite considerar problemas de la forma:

$$\min_{\lambda_g} \frac{1}{2}\lambda_g^T \hat{Y}\hat{K}\hat{Y}\lambda_g - \lambda_g^T e$$

$$\text{sujeto a:} \quad e^T\hat{Y}\lambda_g = 0, \quad 0 \leq \lambda_g \leq C$$

Comentarios.

- La conexión entre nodos de un clúster, la transmisión de datos entre ellos y el manejo de memoria es posible consultar en la referencia de Woodsend y Gondzio.
- La formulación dual de la MSV permite considerar problemas de la forma:

$$\min_{\lambda_g} \frac{1}{2}\lambda_g^T \hat{Y}\hat{K}\hat{Y}\lambda_g - \lambda_g^T e$$

$$\text{sujeto a:} \quad e^T\hat{Y}\lambda_g = 0, \quad 0 \le \lambda_g \le C$$

donde $\hat{K}$ es una matriz con elementos $\hat{K}_{i,j} = \hat{K}(\mathbf{x}_i, \mathbf{x}_j)$ y $\hat{K}(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)$, ϕ no lineal.

Comentarios.

- La conexión entre nodos de un clúster, la transmisión de datos entre ellos y el manejo de memoria es posible consultar en la referencia de Woodsend y Gondzio.
- La formulación dual de la MSV permite considerar problemas de la forma:

$$\min_{\lambda_g} \frac{1}{2}\lambda_g^T \hat{Y}\hat{K}\hat{Y}\lambda_g - \lambda_g^T e$$

$$\text{sujeto a:} \quad e^T \hat{Y}\lambda_g = 0, \quad 0 \leq \lambda_g \leq C$$

  donde $\hat{K}$ es una matriz con elementos $\hat{K}_{i,j} = \hat{K}(\mathbf{x}_i, \mathbf{x}_j)$ y $\hat{K}(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$, ϕ no lineal.
- En la referencia de Woodsend y Gondzio se trabaja este enfoque para realizar en paralelo MSV.