Machines à vecteurs de support

- Idée 1: séparation linéaire avec une marge maximale
  - marge fonctionnelle:

$$\gamma_i = f(\mathbf{x}_i) y_i = (\mathbf{w}^t \mathbf{x}_i + w_0) y_i$$

  - marge géométrique:

$$\gamma_i^{(g)} = \frac{1}{\|\mathbf{w}\|} (\mathbf{w}^t \mathbf{x}_i + w_0) y_i = \frac{1}{\|\mathbf{w}\|} \gamma_i$$

- Optimisations équivalentes
  - maximiser la marge géométrique
  - maximiser la marge fonctionnelle sous la contrainte $\|\mathbf{w}\| = 1$
  - minimiser $\|\mathbf{w}\|$ sous la contrainte $\gamma_i \geq 1$

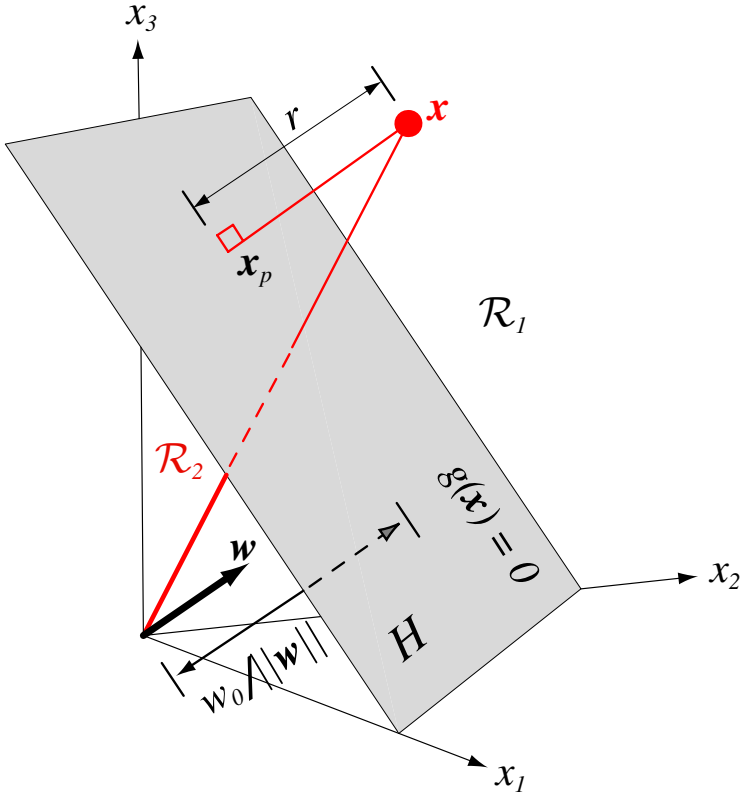
Machines à vecteurs de support

- Géométrie – deux classes
  - $r =$ distance algébrique de $\mathbf{x}$ et H:

$$\begin{aligned} \mathbf{x} &= \mathbf{x}_p + r\frac{\mathbf{w}}{\|\mathbf{w}\|} \\ g(\mathbf{x}) &= \mathbf{w}^t\mathbf{x} + w_0 = r\|\mathbf{w}\| \\ r &= \frac{g(\mathbf{x})}{\|\mathbf{w}\|} \end{aligned}$$

Machines à vecteurs de support

- Géométrie – deux classes



Machines à vecteurs de support

- Problème d'optimisation:
  - soit $D_n = \big((\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\big)$ linéairement séparable
  - minimiser $\|\mathbf{w}\|^2 = \mathbf{w}^t\mathbf{w}$
  - sous les contraintes $\gamma_i = (\mathbf{w}^t\mathbf{x}_i + w_0)y_i \geq 1, \;\; i = 1, \ldots, n$

- Résultat:
  - l'hyperplan ($\mathbf{w}^t\mathbf{x} + w_0 = 0$) avec une marge géométrique $\dfrac{1}{\|\mathbf{w}\|}$ maximale

Machines à vecteurs de support

- Problème primal – théorème de Kuhn-Tucker
  - minimiser par rapport à $\mathbf{w}$ et w_0 et maximiser par rapport à α:

$$\begin{aligned} L(\mathbf{w}, w_0, \alpha) &= \frac{1}{2}\mathbf{w}^t\mathbf{w} - \sum_{i=1}^{n} \alpha_i(\gamma_i - 1) \\ &= \frac{1}{2}\mathbf{w}^t\mathbf{w} - \sum_{i=1}^{n} \alpha_i\big((\mathbf{w}^t\mathbf{x}_i + w_0)y_i - 1\big) \end{aligned}$$

  - sous les contraintes $\alpha_i \geq 0, \;\; i = 1, \ldots, n$

Machines à vecteurs de support

- Optimisation:
  - les gradients:

$$\frac{\partial L(\mathbf{w}, w_0, \alpha)}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i = \mathbf{0}$$

$$\frac{\partial L(\mathbf{w}, w_0, \alpha)}{\partial w_0} = \sum_{i=1}^{n} \alpha_i y_i = 0$$

  - resubstitution: maximiser par rapport à α (problème dual):

$$W(\alpha) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^t \mathbf{x}_j - \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^t \mathbf{x}_j + \sum_{i=1}^{n} \alpha_i$$

$$= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^t \mathbf{x}_j$$

  - sous les contraintes $\alpha_i \geq 0, \ \ i = 1, \ldots, n$ et $\sum_{i=1}^{n} \alpha_i y_i = 0$

Machines à vecteurs de support

- La solution
  - $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i^* y_i \mathbf{x}_i$
  - $w_0^* = -\frac{1}{2}\left(\max_{y_i=-1} \mathbf{w}^* \mathbf{x}_i + \min_{y_i=1} \mathbf{w}^* \mathbf{x}_i \right)$

Machines à vecteurs de support

- La structure de la solution
  - $\mathbf{w}^*$ est une combinaison linéaire des points d'entraînement
  - théorème de Kuhn-Tucker:

$$\alpha_i^* \left((\mathbf{w}^{*t} \mathbf{x}_i + w_0^*) y_i - 1 \right) = 0, \;\; i = 1, \ldots, n$$

  - si $\gamma_i > 1$ alors $\alpha_i^* = 0$
  - si $\alpha_i^* > 0$ alors $\gamma_i = 1$: vecteurs de support
  - $\mathbf{w}^* = \sum_{i \in sv} \alpha_i^* y_i \mathbf{x}_i$
  - $f^*(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i^* y_i \mathbf{x}_i^t \mathbf{x} + w_0^* = \sum_{i \in sv} \alpha_i^* y_i \mathbf{x}_i^t \mathbf{x} + w_0^*$

Machines à vecteurs de support

- La structure de la solution
  - pour $j \in sv$

$$\gamma_j = y_j f^*(\mathbf{x}_j) = y_j \left(\sum_{i \in sv} \alpha_i^* y_i \mathbf{x}_i^t \mathbf{x}_j + w_0^* \right) = 1$$

  - alors

$$\begin{aligned} \mathbf{w}^{*t}\mathbf{w}^* &= \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i^* \alpha_j^* y_i y_j \mathbf{x}_i^t \mathbf{x}_j = \sum_{j \in sv} \alpha_j^* y_j \sum_{i \in sv} \alpha_i^* y_i \mathbf{x}_i^t \mathbf{x}_j = \sum_{j \in sv} \alpha_j^* (1 - y_j w_0^*) \\ &= \sum_{j \in sv} \alpha_j^* \end{aligned}$$

  - la marge maximale:

$$\gamma^* = \frac{1}{\|\mathbf{w}\|} = \left(\sum_{i \in sv} \alpha_i^* \right)^{-1/2}$$

Machines à vecteurs de support

- Idée 2: le noyau
  - l'optimisation: $W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^t \mathbf{x}_j$
  - la fonction optimale: $f^*(\mathbf{x}) = \sum_{i \in sv} \alpha_i^* y_i \mathbf{x}_i^t \mathbf{x} + w_0^*$
- Remplacer $\mathbf{x}^t \mathbf{x}'$ par une fonction de noyaux $K(\mathbf{x}, \mathbf{x}')$
  - l'optimisation: $W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$
  - la fonction optimale: $f^*(\mathbf{x}) = \sum_{i \in sv} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + w_0^*$
  - matrice de Gram: $\mathbf{G}_{i,j} = K(\mathbf{x}_i, \mathbf{x}_j)$
  - équivalent à une transformation non-linéaire dans une espace de traits de dimension très élévée

Machines à vecteurs de support

- Exemples de noyaux
  - linéaire: $K^{\ell}(\mathbf{x},\mathbf{x}') = \mathbf{x}^t\mathbf{x}'$
  - polynômial: $K^{p}_{d}(\mathbf{x},\mathbf{x}') = (\mathbf{x}^t\mathbf{x}'+1)^d$
  - base radiale: $K^{r}(\mathbf{x},\mathbf{x}') = K(\|\mathbf{x}-\mathbf{x}'\|)$
    - gaussien: $K^{g}_{\gamma}(\mathbf{x},\mathbf{x}') = e^{-\gamma\|\mathbf{x}-\mathbf{x}'\|^2}$
  - réseaux de neurones: $K^{\sigma}(\mathbf{x},\mathbf{x}') = \sigma(s\mathbf{x}^t\mathbf{x}'+c)$
    - sigmoïde: $K^{t}_{s,c}(\mathbf{x},\mathbf{x}') = \tanh(s\mathbf{x}^t\mathbf{x}'+c)$

Machines à vecteurs de support

- Discrimination linéaire généralisée
  - linéaire:

$$f^*(\mathbf{x}) = \mathbf{w}^{*t}\mathbf{x} + w_0^* = \sum_{j=1}^{d} w_j^* x^{(j)} + w_0^* = \sum_{i \in sv} \alpha_i^* y_i \mathbf{x}_i^t \mathbf{x} + w_0^*$$

  - généralisée:

$$f^*(\mathbf{x}) = \sum_{j=1}^{D} w_j^* \phi^{(j)}(\mathbf{x}) + w_0^* = \sum_{i \in sv} \alpha_i^* y_i \sum_{j=1}^{D} \phi^{(j)}(\mathbf{x}_i)\phi^{(j)}(\mathbf{x}) + w_0^*$$

  - produit scalaire dans l'espace de traits:

$$\phi^t(\mathbf{x})\phi(\mathbf{x}') = \sum_{j=1}^{D} \phi^{(j)}(\mathbf{x})\phi^{(j)}(\mathbf{x}')$$

Machines à vecteurs de support

- Discrimination linéaire généralisée
  - exemple:

$$\begin{aligned}
\phi^{(j)}(\mathbf{x}) &= \sqrt{2}x^{(j)}, \quad j = 1, \ldots, d, \\
\phi^{(id+j)}(\mathbf{x}) &= x^{(i)}x^{(j)}, \quad i, j = 1, \ldots, d \\
\phi^{(d^2)}(\mathbf{x}) &= 1
\end{aligned}$$

  - produit scalaire dans l'espace des traits:

$$\begin{aligned}
\phi^t(\mathbf{x})\phi(\mathbf{x}') &= \sum_{j=1}^{d^2} \phi^{(j)}(\mathbf{x})\phi^{(j)}(\mathbf{x}') \\
&= \sum_{j=1}^{d} \sqrt{2}x^{(j)}\sqrt{2}x'^{(j)} + \sum_{i=1}^{d}\sum_{j=1}^{d} x^{(i)}x^{(j)}x'^{(i)}x'^{(j)} + 1 \\
&= 2\mathbf{x}^t\mathbf{x}' + (\mathbf{x}^t\mathbf{x}')^2 + 1 \\
&= (\mathbf{x}^t\mathbf{x}' + 1)^2 = K_2^p(\mathbf{x}, \mathbf{x}')
\end{aligned}$$

Machines à vecteurs de support

- Caractérisation des noyaux
  - $K(\mathbf{x},\mathbf{x}') = \phi^t(\mathbf{x})\phi(\mathbf{x}') = \sum_{j=1}^{D} \phi^{(j)}(\mathbf{x})\phi^{(j)}(\mathbf{x}')$
  - commutativité: $K(\mathbf{x},\mathbf{x}') = K(\mathbf{x}',\mathbf{x})$
  - inégalité de Cauchy-Schwartz:

$$\begin{aligned} K(\mathbf{x},\mathbf{x}')^2 &= \left(\phi^t(\mathbf{x})\phi(\mathbf{x}')\right)^2 \\ &\leq \|\phi(\mathbf{x})\|^2 \|\phi(\mathbf{x}')\|^2 \\ &= \phi^t(\mathbf{x})\phi(\mathbf{x}) \cdot \phi^t(\mathbf{x}')\phi(\mathbf{x}') \\ &= K(\mathbf{x},\mathbf{x})K(\mathbf{x}',\mathbf{x}') \end{aligned}$$

Machines à vecteurs de support

- Théorème de Mercer:
  - $\phi^t(\mathbf{x})\phi(\mathbf{x}') = K(\mathbf{x}, \mathbf{x}')$
- Conditions
  - diagonaliser la matrice de Gram: $\mathbf{G}_{i,j} = K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{V}\Lambda\mathbf{V}$
  - valeurs propres: λ_t, vecteurs propres: $\mathbf{v}_t$
  - condition suffisante: $\mathbf{G}$ est positif semi-definit ($\forall t : \lambda_t > 0$) pour tout $\mathcal{X}_n = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$
  - D peut être ∞!!!

Machines à vecteurs de support

- Idée 3: les variables d'écart (slack variables)
  - cas non-séparable
- Permettre des erreurs:
  - minimiser $\|\mathbf{w}\|^2 = \mathbf{w}^t\mathbf{w}$
  - sous les contraintes $\gamma_i = (\mathbf{w}^t\mathbf{x}_i + w_0)y_i \geq 1 - \xi_i, \;\; i = 1, \ldots, n$
  - où $\xi_i \geq 0, \;\; i = 1, \ldots, n$
  - minimiser l'erreur $\equiv$ minimiser le nombre de points avec $\xi_i > 0$: NP-difficile

Machines à vecteurs de support

- Problème soluble 1: minimiser $\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i^2$
  - on peut supprimer la contrainte de positivité des ξ_i
  - C est un hyper-paramètre réglé par la validation croisée (par exemple)

Machines à vecteurs de support

- Problème primal
  - minimiser par rapport à $\mathbf{w}$, ξ et w_0 et maximiser par rapport à α:

$$L(\mathbf{w}, w_0, \xi, \alpha) = \frac{1}{2}\mathbf{w}^t\mathbf{w} + \frac{C}{2}\sum_{i=1}^{n}\xi_i^2 - \sum_{i=1}^{n}\alpha_i\left[(\mathbf{w}^t\mathbf{x}_i + w_0)y_i - 1 + \xi_i\right]$$

  - sous les contraintes $\alpha_i \geq 0, \;\; i = 1, \ldots, n$
  - les gradients:

$$\begin{aligned}
\frac{\partial L(\mathbf{w}, w_0, \alpha)}{\partial \mathbf{w}} &= \mathbf{w} - \sum_{i=1}^{n}\alpha_i\mathbf{x}_i y_i = \mathbf{0} \\
\frac{\partial L(\mathbf{w}, w_0, \alpha)}{\partial \xi} &= C\xi - \alpha = \mathbf{0} \\
\frac{\partial L(\mathbf{w}, w_0, \alpha)}{\partial w_0} &= \sum_{i=1}^{n}\alpha_i y_i = 0
\end{aligned}$$

Machines à vecteurs de support

- Problème dual
  - resubstitution: maximiser par rapport à α:

$$\begin{aligned} W(\alpha) &= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) + \frac{1}{2C} \alpha^t \alpha - \frac{1}{C} \alpha^t \alpha \\ &= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) - \frac{1}{2C} \alpha^t \alpha \\ &= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \left(K(\mathbf{x}_i, \mathbf{x}_j) + \frac{1}{C} \delta_{i,j} \right) \end{aligned}$$

  - sous les contraintes $\alpha_i \geq 0, \;\; i = 1, \ldots, n$ et $\sum_{i=1}^{n} \alpha_i y_i = 0$

Machines à vecteurs de support

- La solution
  - $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i^* y_i \mathbf{x}_i = \sum_{i \in sv} \alpha_i^* y_i \mathbf{x}_i$
  - $f^*(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + w_0^* = \sum_{i \in sv} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + w_0^*$
  - w_0^* est choisi tel que $\gamma_i = f(\mathbf{x}_i) y_i \geq 1 - \xi_i = 1 - \frac{\alpha_i^*}{C}$
  - la marge obtenue: $\gamma = \left(\sum_{i \in sv} \alpha_i^* - \frac{1}{C} \alpha^{*t} \alpha^* \right)^{-1/2}$

Machines à vecteurs de support

- Problème soluble 2: minimiser $\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i$

- Problème primal

  - minimiser par rapport à $\mathbf{w}$, ξ et w_0 et maximiser par rapport à α:

$$L(\mathbf{w}, w_0, \xi, \alpha, \mathbf{r}) = \frac{1}{2}\mathbf{w}^t\mathbf{w} + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\left[(\mathbf{w}^t\mathbf{x}_i + w_0)y_i - 1 + \xi_i\right] - \sum_{i=1}^{n} r_i\xi_i$$

  - sous les contraintes $\alpha_i \geq 0, r_i \geq 0 \;\; i = 1, \ldots, n$

Machines à vecteurs de support

- Problème dual
  - les gradients:

$$\frac{\partial L(\mathbf{w}, w_0, \alpha)}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i = \mathbf{0}$$

$$\frac{\partial L(\mathbf{w}, w_0, \alpha)}{\partial \xi} = C - \alpha_i - r_i = 0$$

$$\frac{\partial L(\mathbf{w}, w_0, \alpha)}{\partial w_0} = \sum_{i=1}^{n} \alpha_i y_i = 0$$

Machines à vecteurs de support

- Problème dual
  - resubstitution: maximiser par rapport à α:

$$W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$$

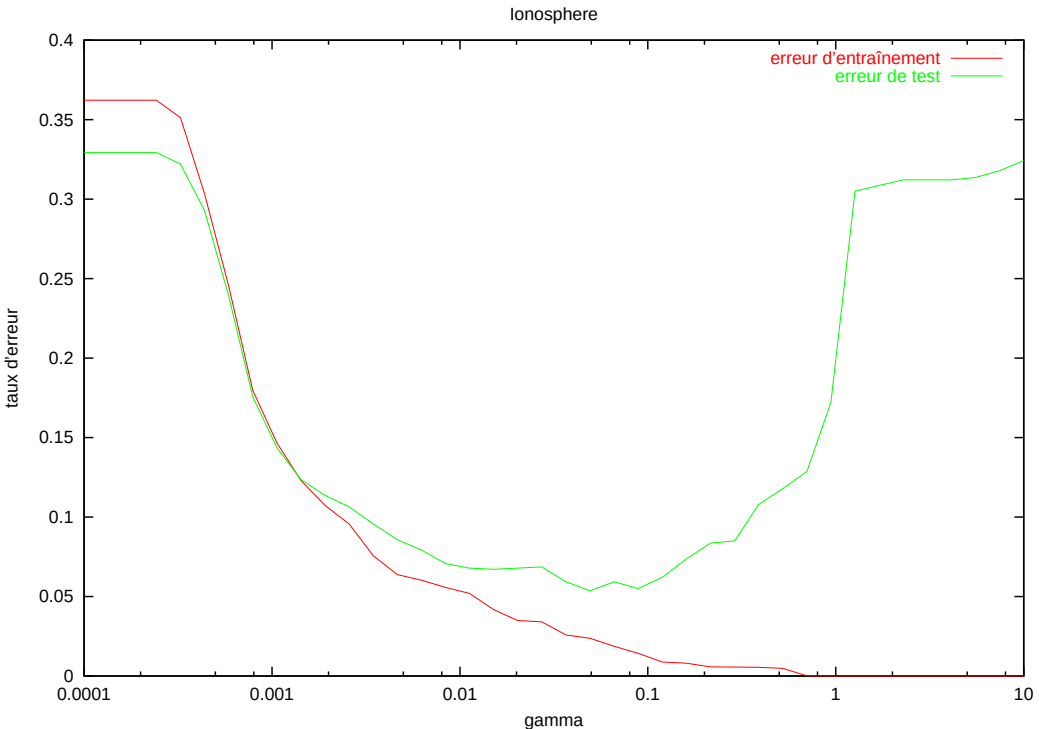
  - sous les contraintes $\alpha_i \geq 0, r_i \geq 0, C - \alpha_i - r_i = 0 \;\; i = 1, \ldots, n$ et $\sum_{i=1}^{n} \alpha_i y_i = 0$
  - $\equiv$ sous les contraintes $0 \leq \alpha_i \leq C$ (contraintes de boîte)

Machines à vecteurs de support

- La solution
  - $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i^* y_i \mathbf{x}_i = \sum_{i \in sv} \alpha_i^* y_i \mathbf{x}_i$
  - $f^*(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + w_0^* = \sum_{i \in sv} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + w_0^*$
  - w_0^* est choisi tel que $\gamma_i = f(\mathbf{x}_i) y_i = 1$ pour tous $i : 0 < \alpha^* < C$
  - la marge obtenue: $\gamma = \left(\sum_{i,j \in sv} \alpha_i^* \alpha_j^* y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \right)^{-1/2}$

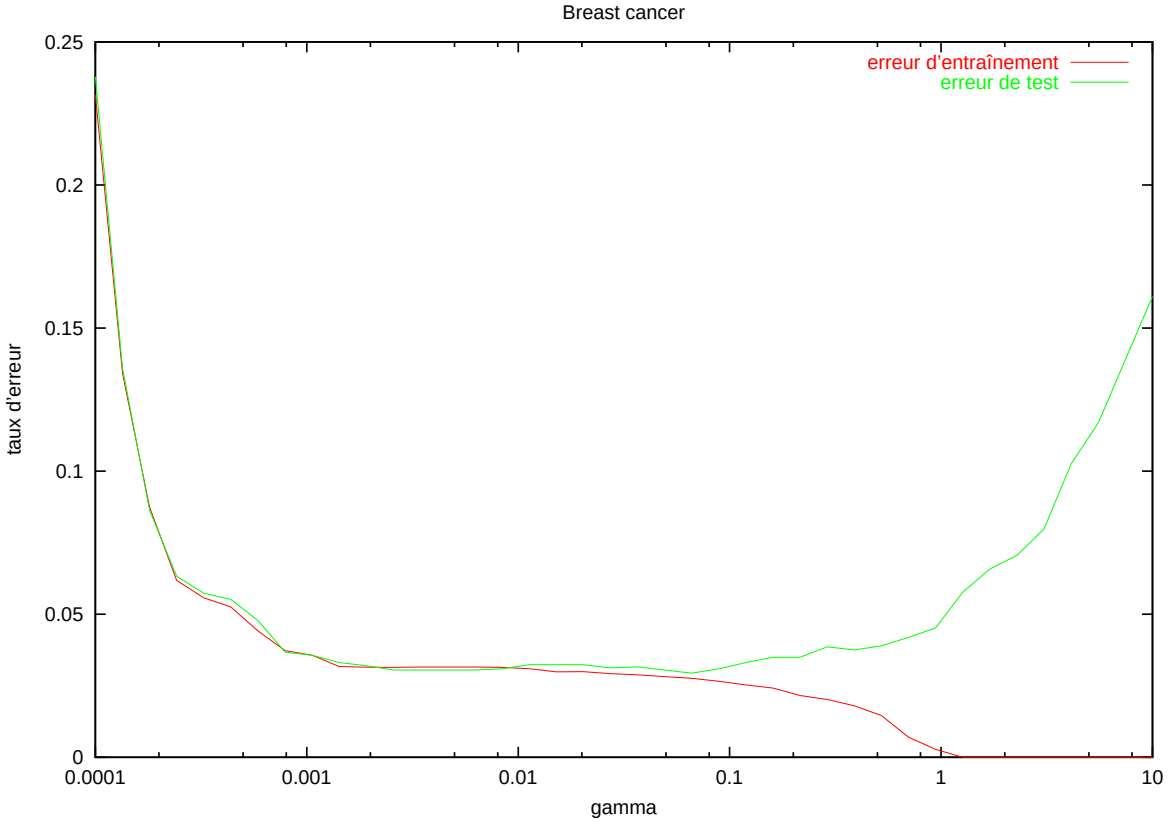
Machines à vecteurs de support

- Expériences
  - 4 données de UCI, test croisé en 10 blocs, contrainte de L_1, $C = 1$, noyau gaussien



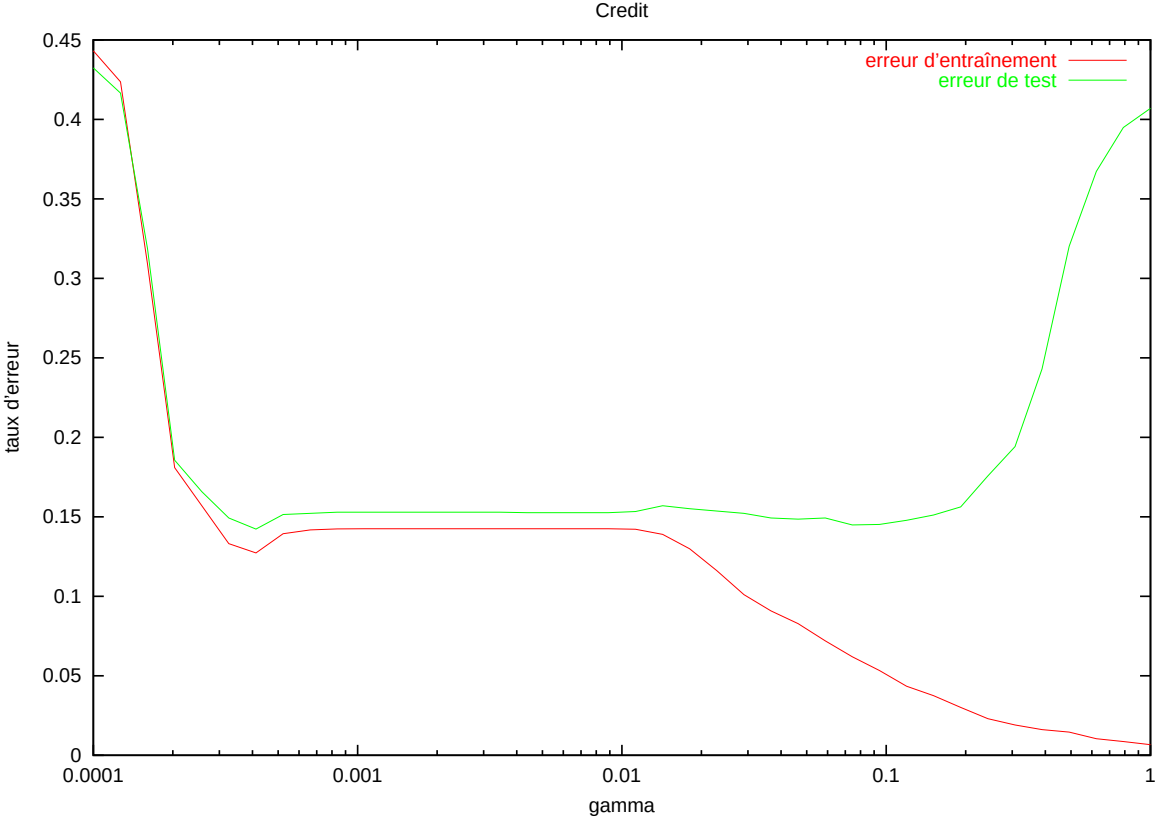
Machines à vecteurs de support

- Expériences



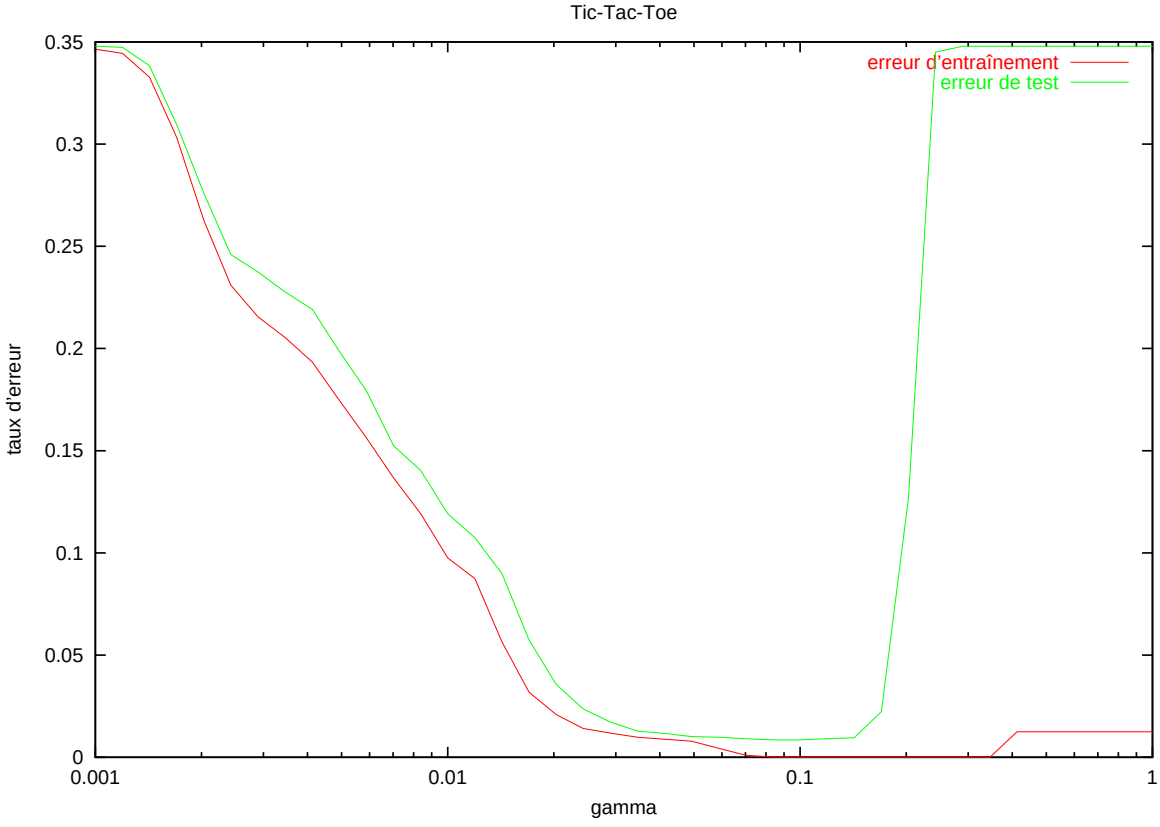
Machines à vecteurs de support

- Expériences



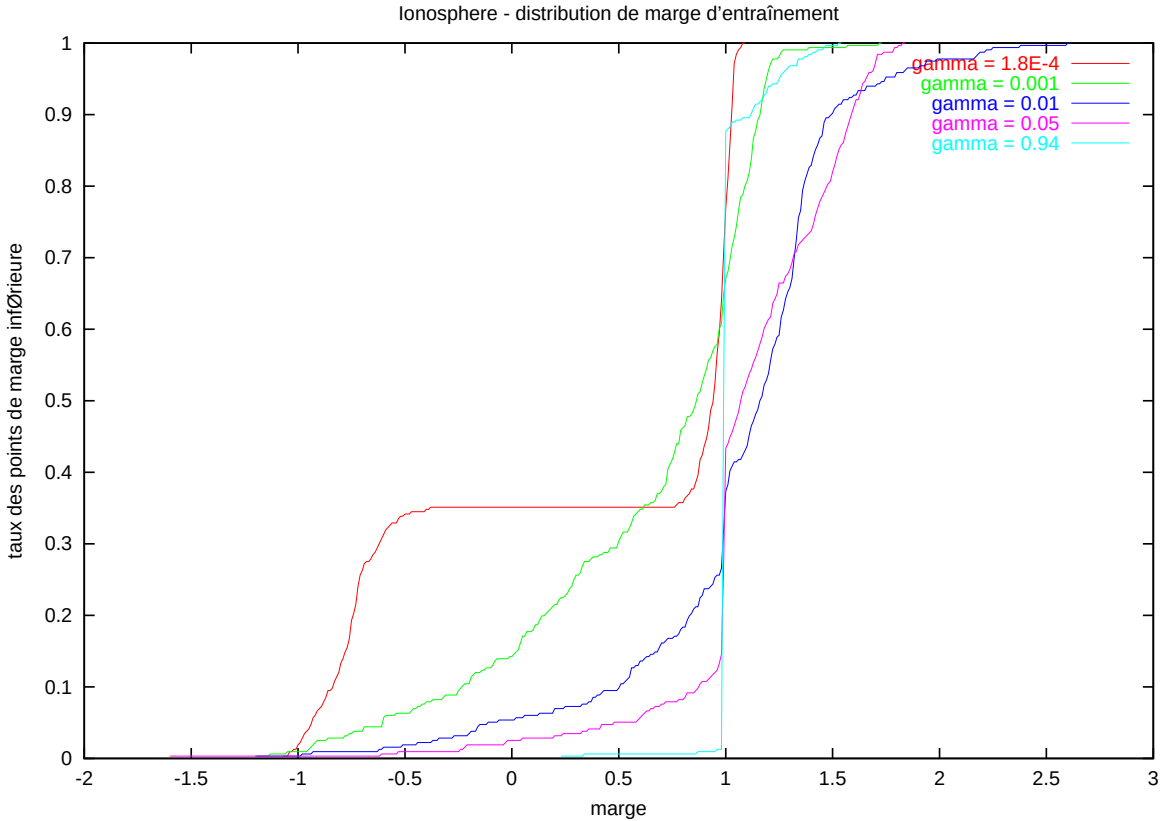
Machines à vecteurs de support

- Expériences



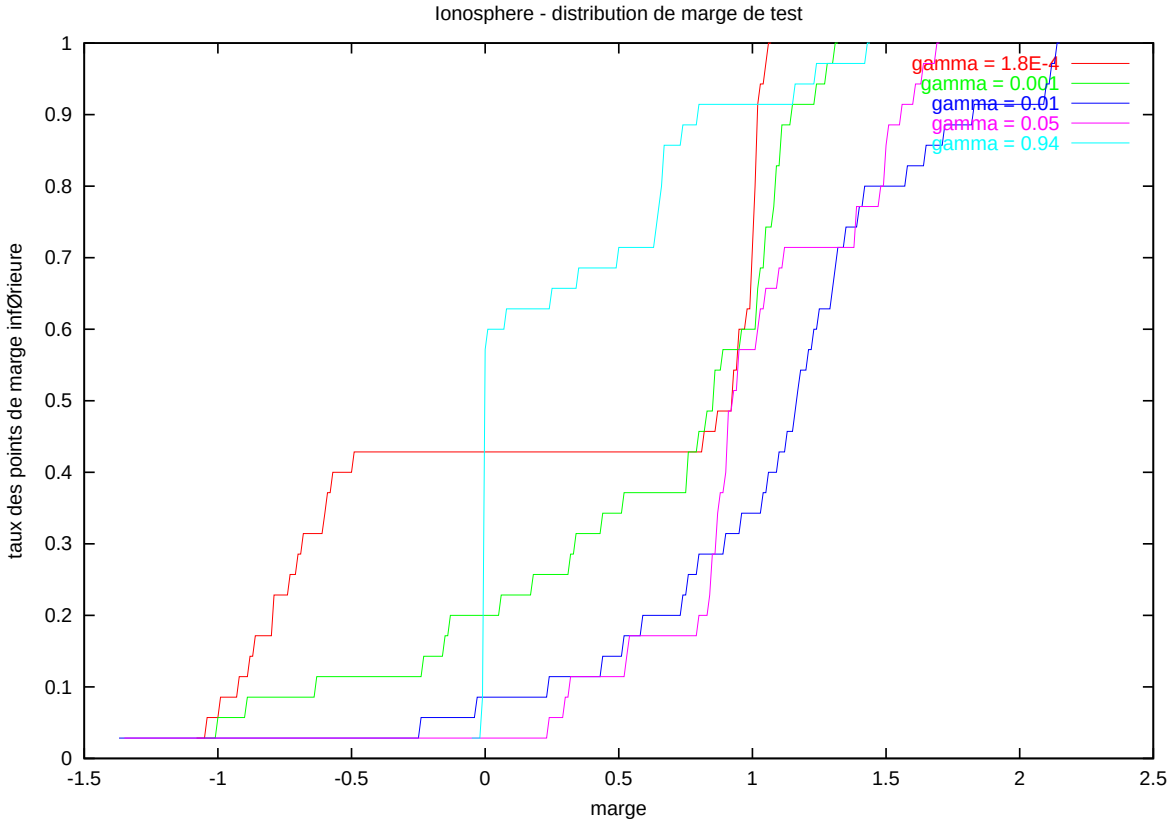
Machines à vecteurs de support

- Distribution de la marge d'entraî nement



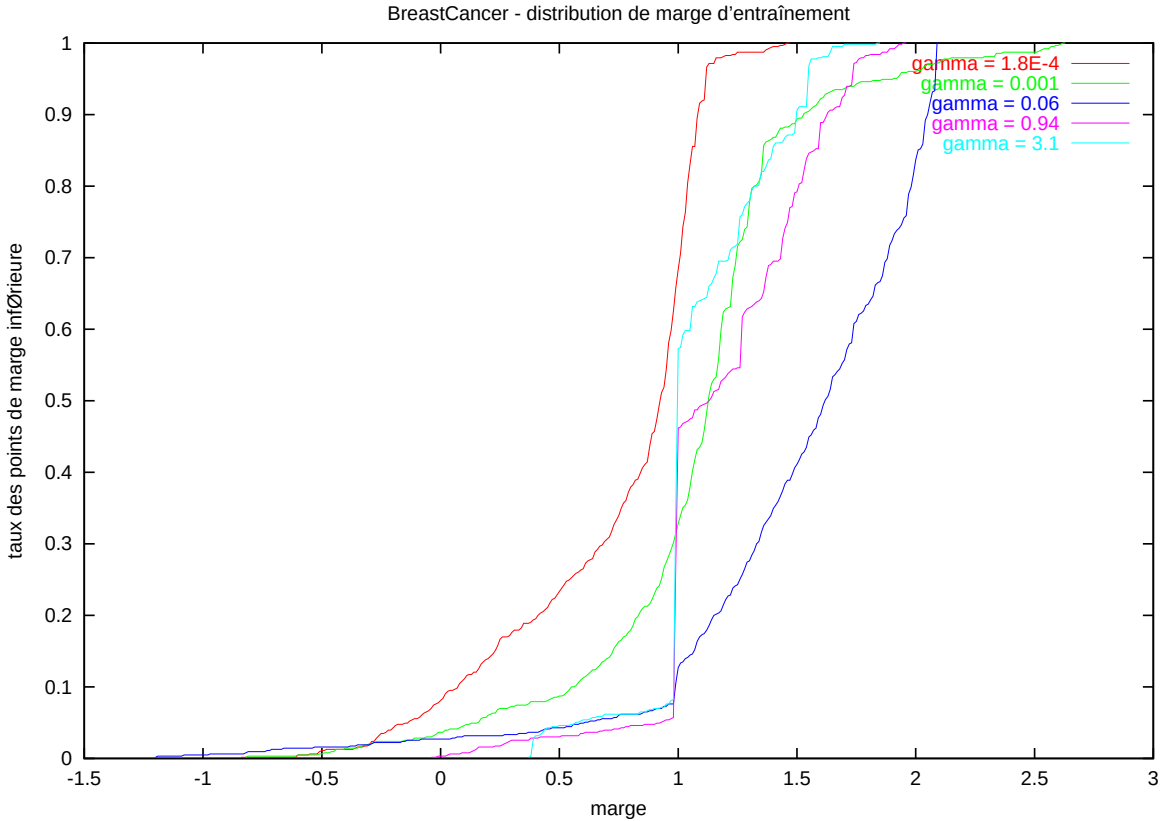
Machines à vecteurs de support

- Distribution de la marge de test



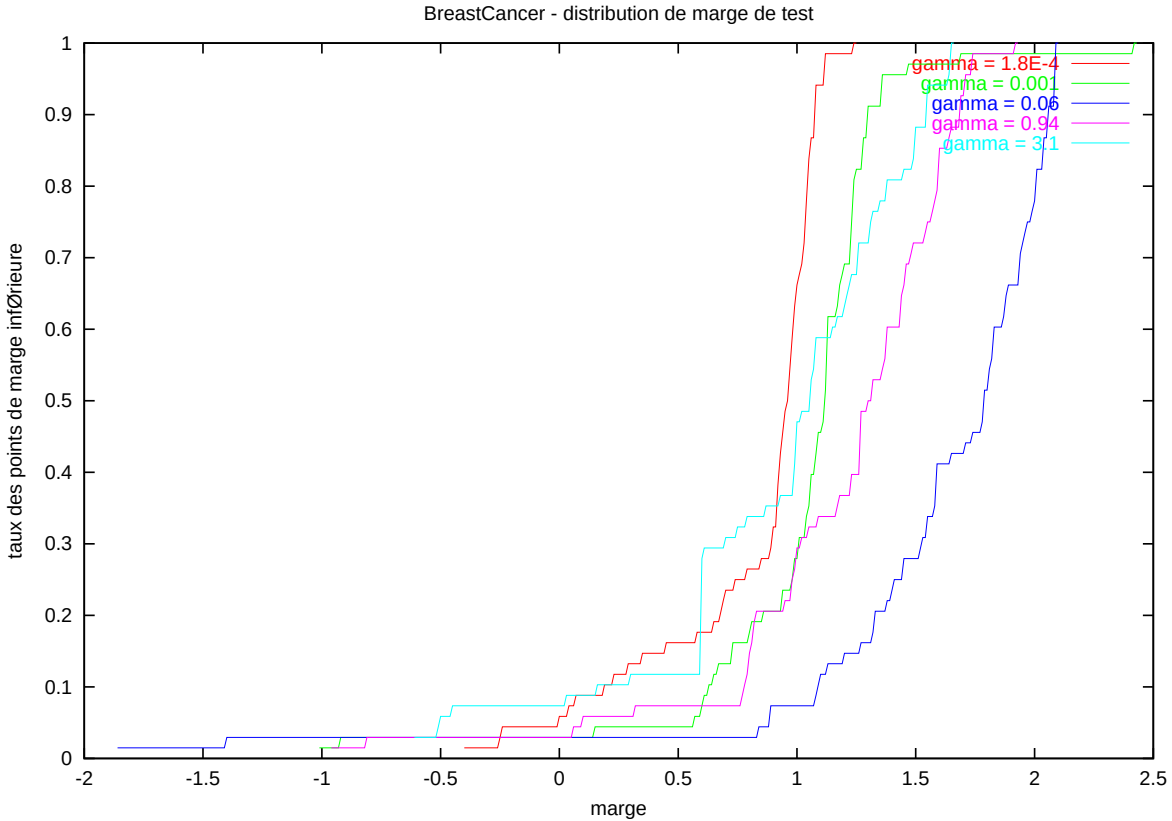
Machines à vecteurs de support

- Distribution de la marge d'entraî nement



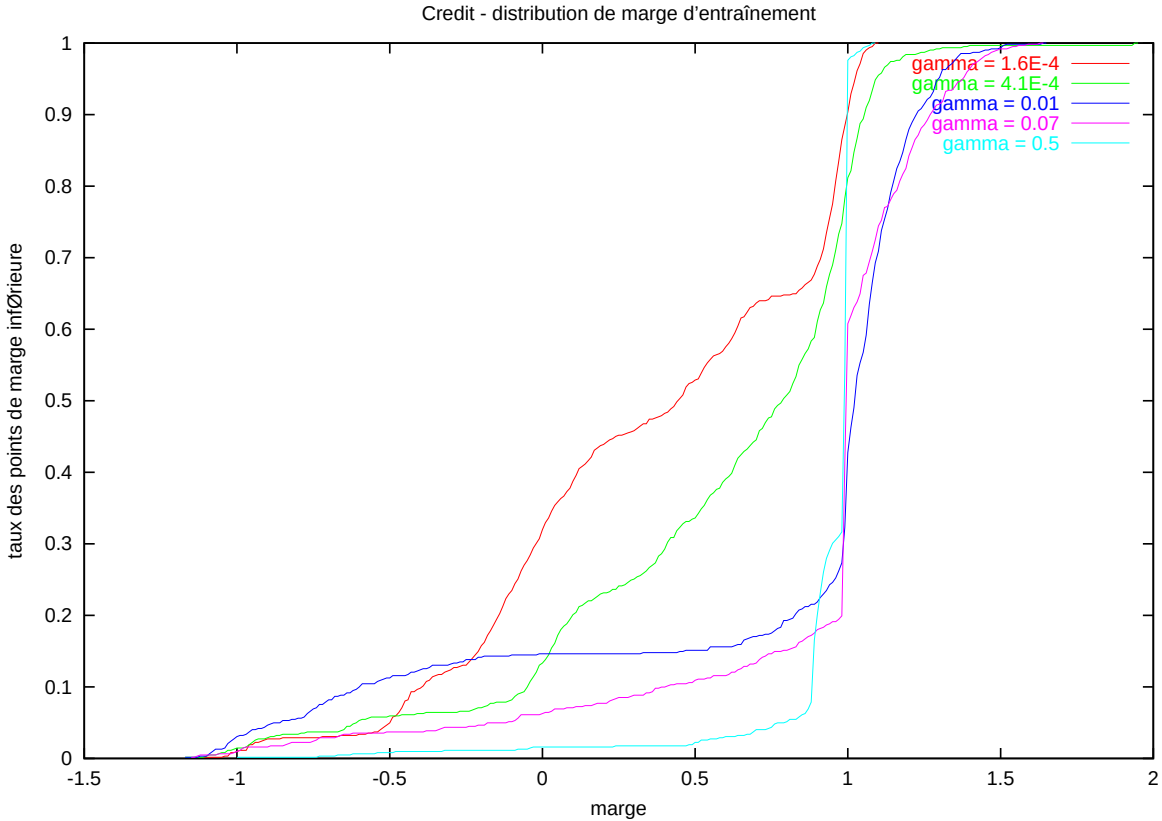
Machines à vecteurs de support

- Distribution de la marge de test



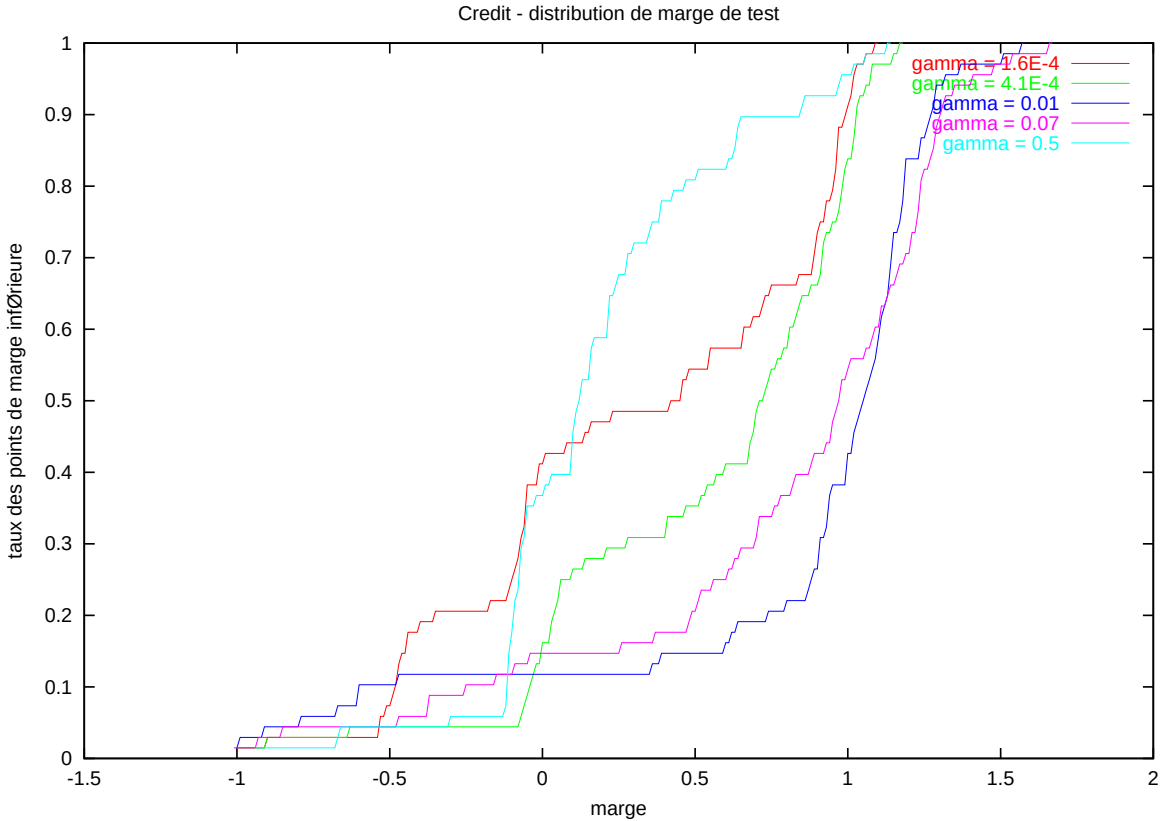
Machines à vecteurs de support

- Distribution de la marge d'entraî nement



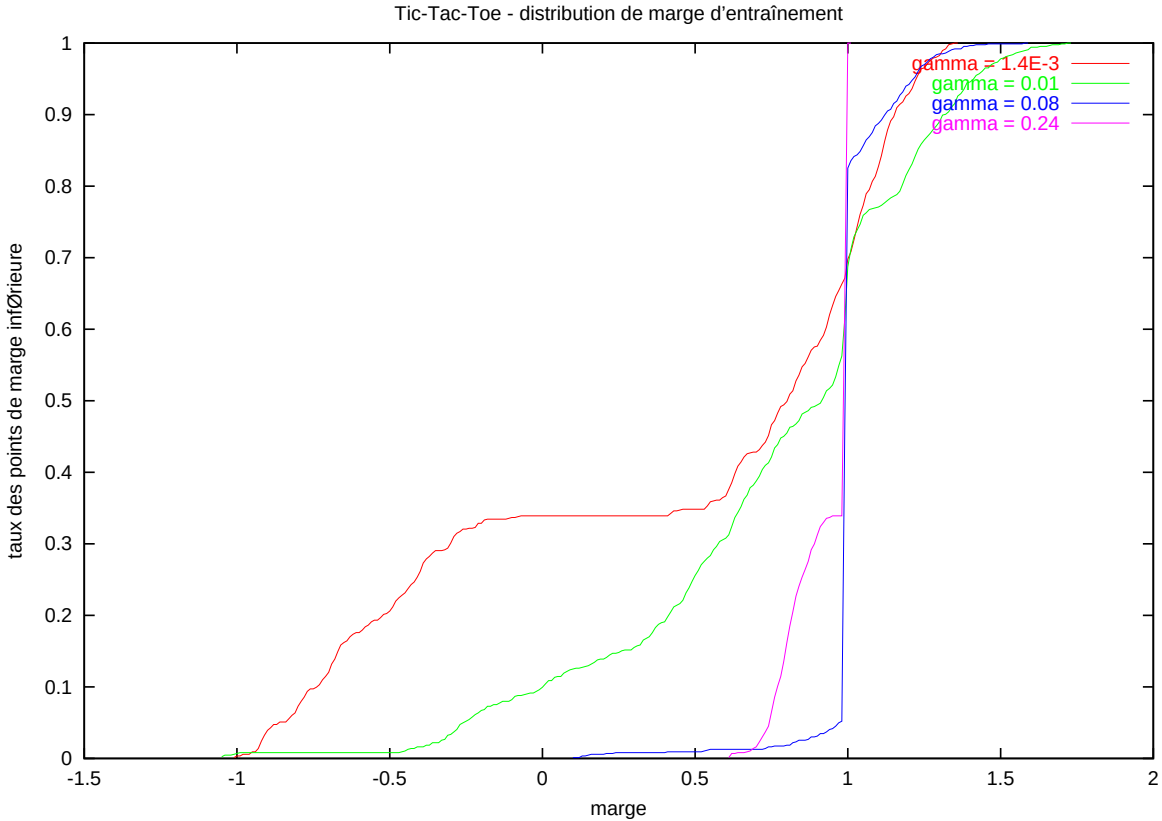
Machines à vecteurs de support

- Distribution de la marge de test



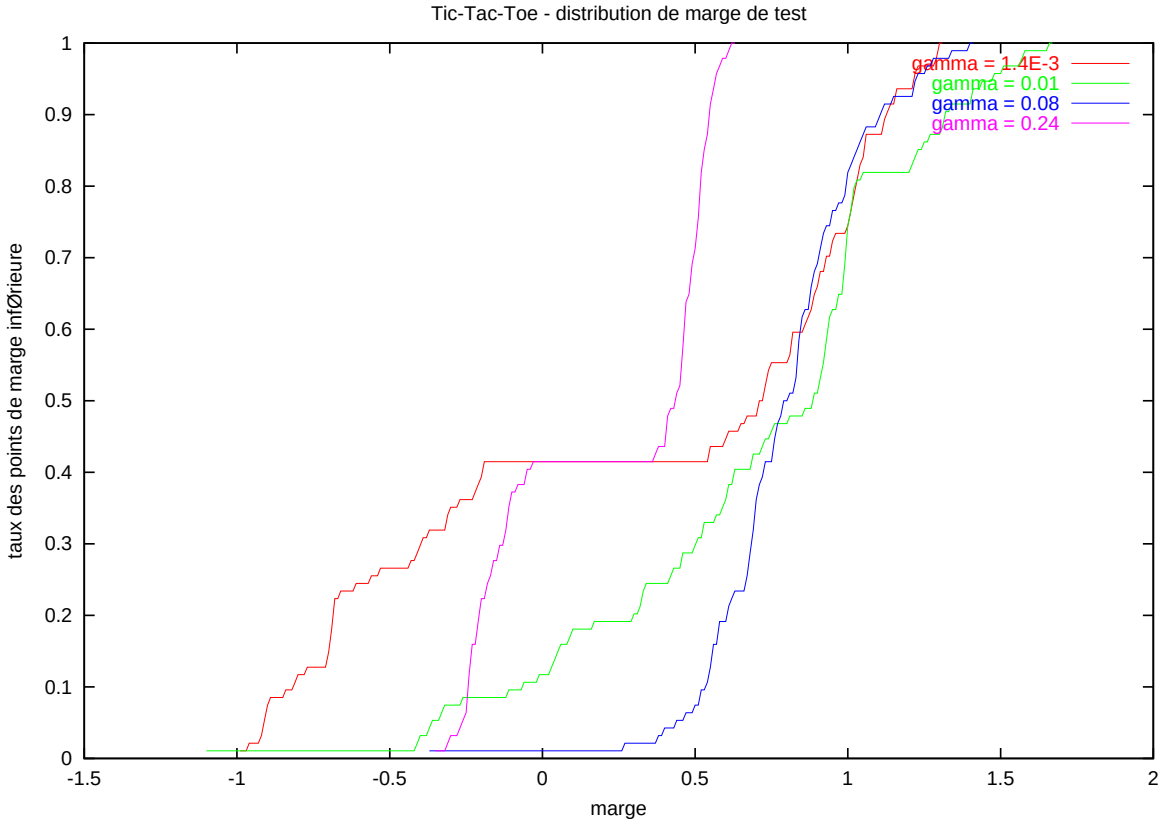
Machines à vecteurs de support

- Distribution de la marge d'entraî nement



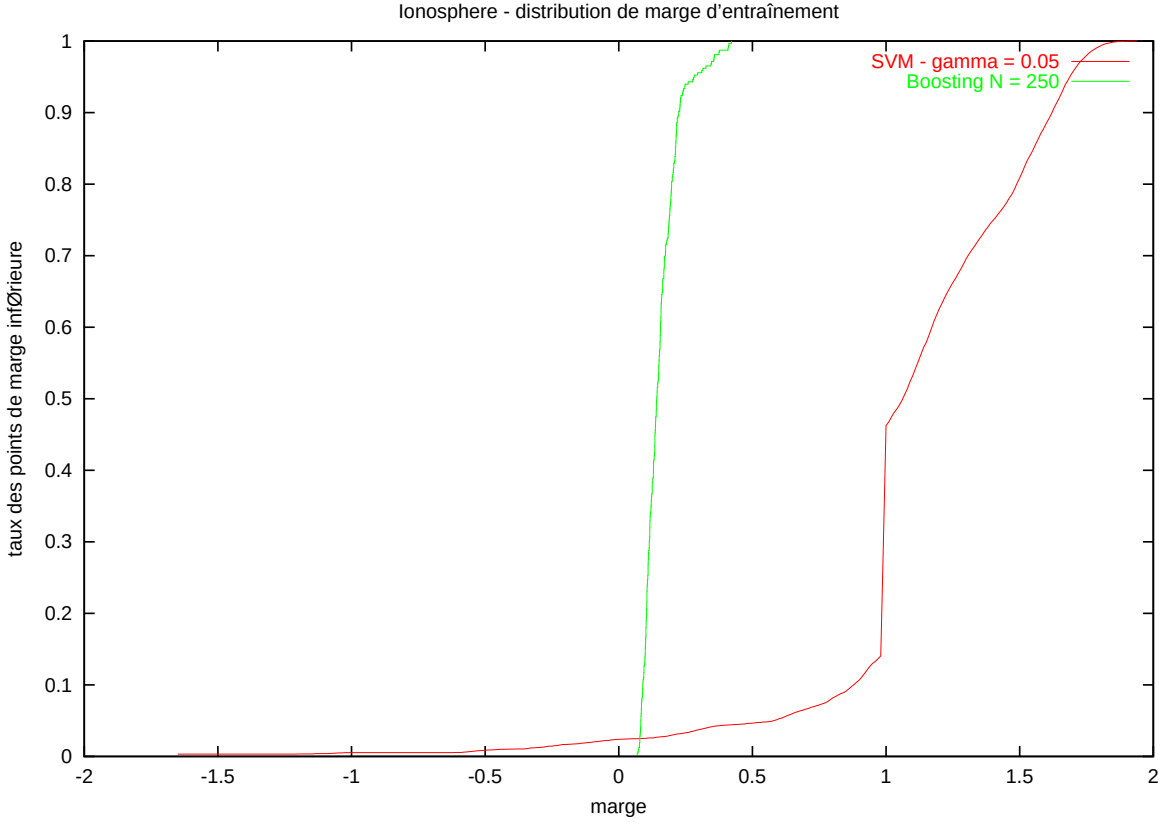
Machines à vecteurs de support

- Distribution de la marge de test



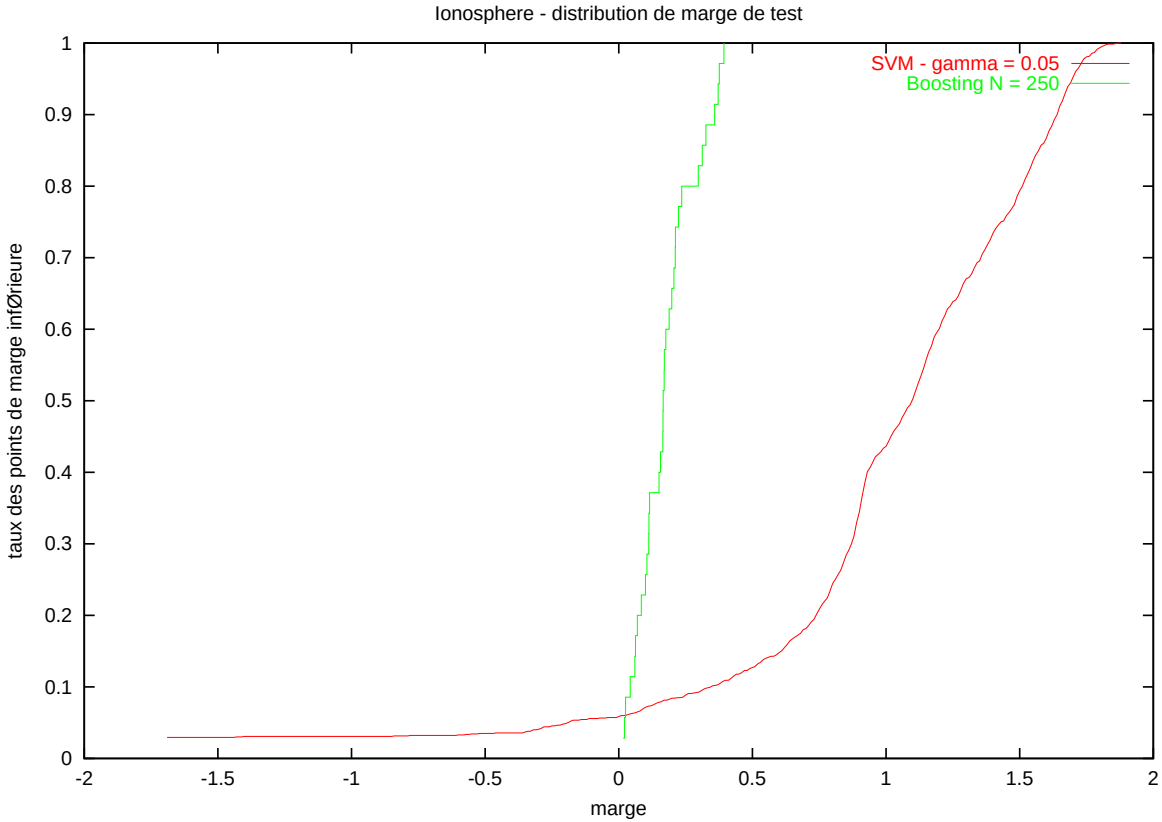
Machines à vecteurs de support

- Comparaison avec boosting (marge d'entraî nement)



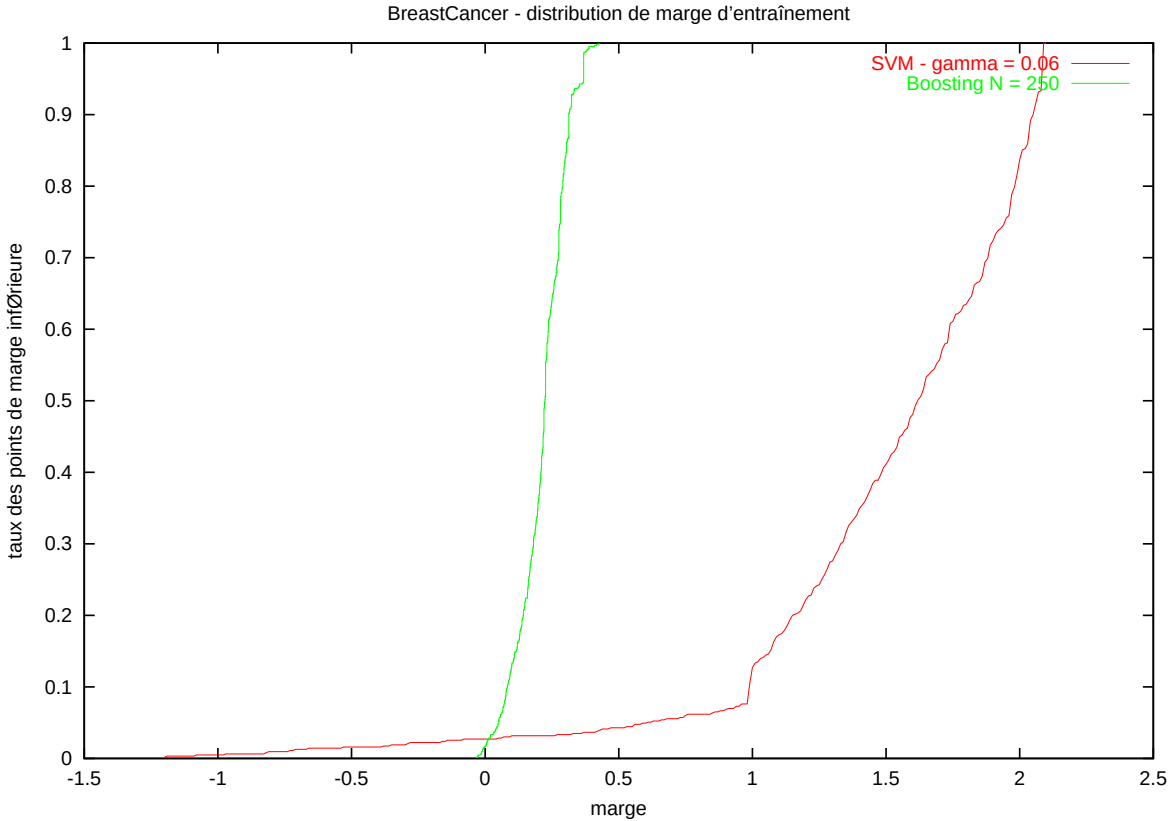
Machines à vecteurs de support

- Comparaison avec boosting (marge de test)



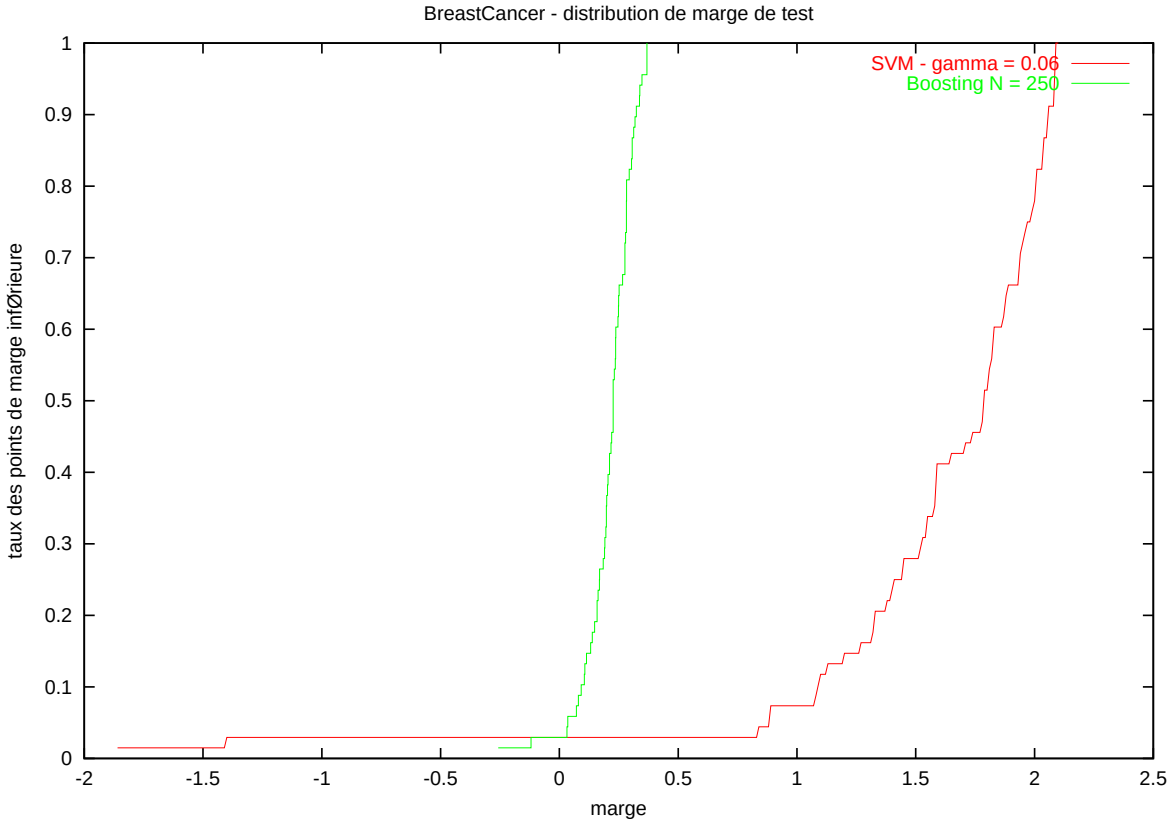
Machines à vecteurs de support

- Comparaison avec boosting (marge d'entraî nement)



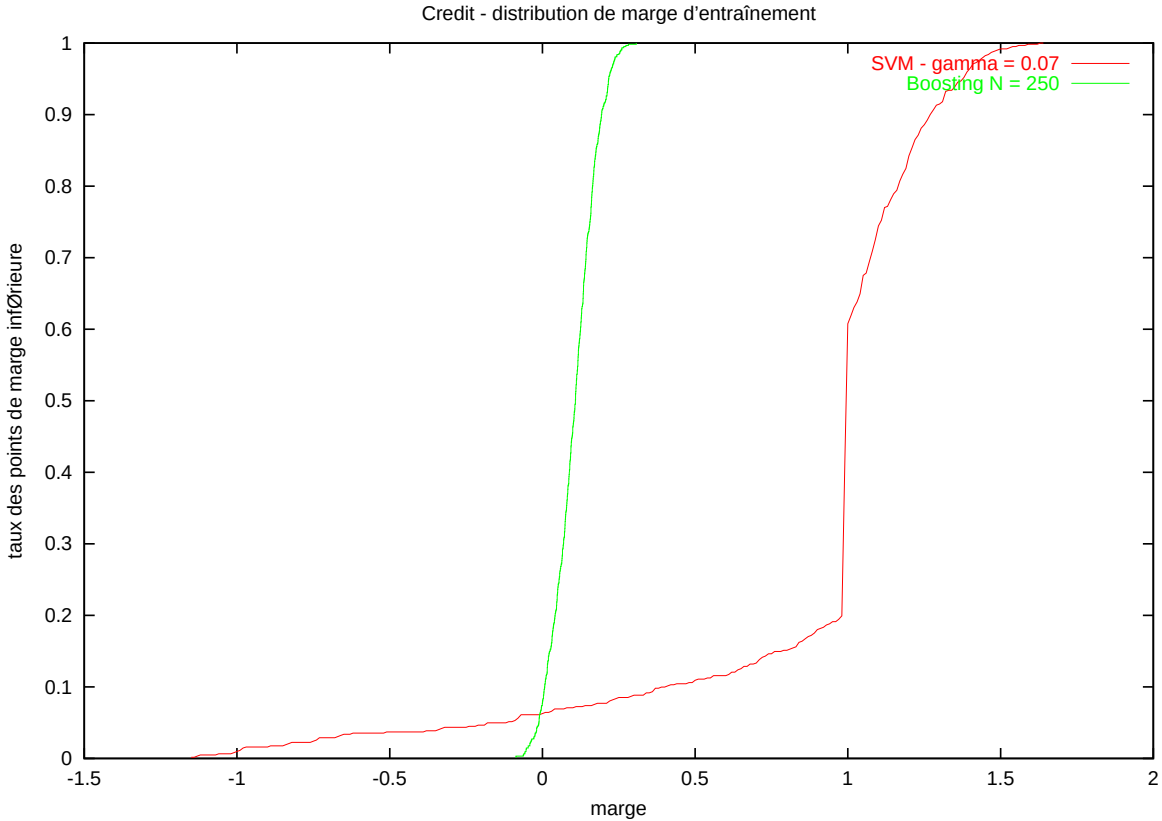
Machines à vecteurs de support

- Comparaison avec boosting (marge de test)



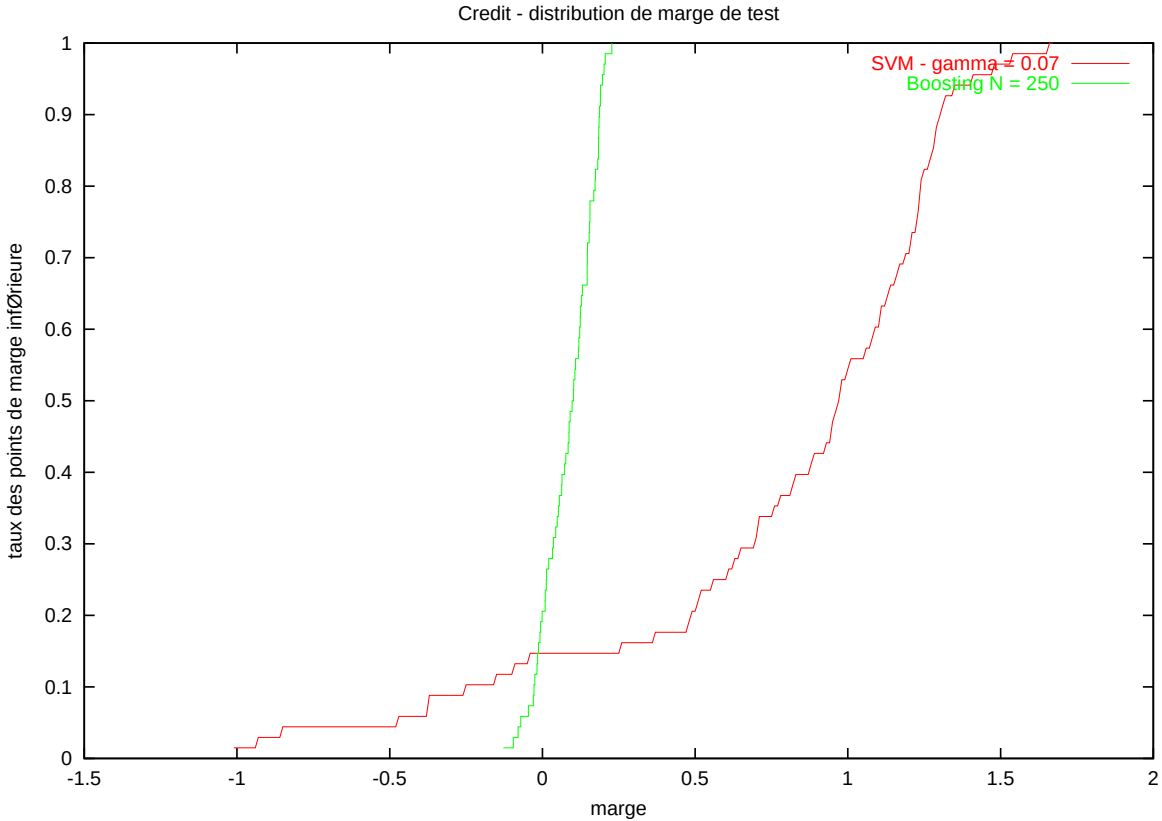
Machines à vecteurs de support

- Comparaison avec boosting (marge d'entraî nement)



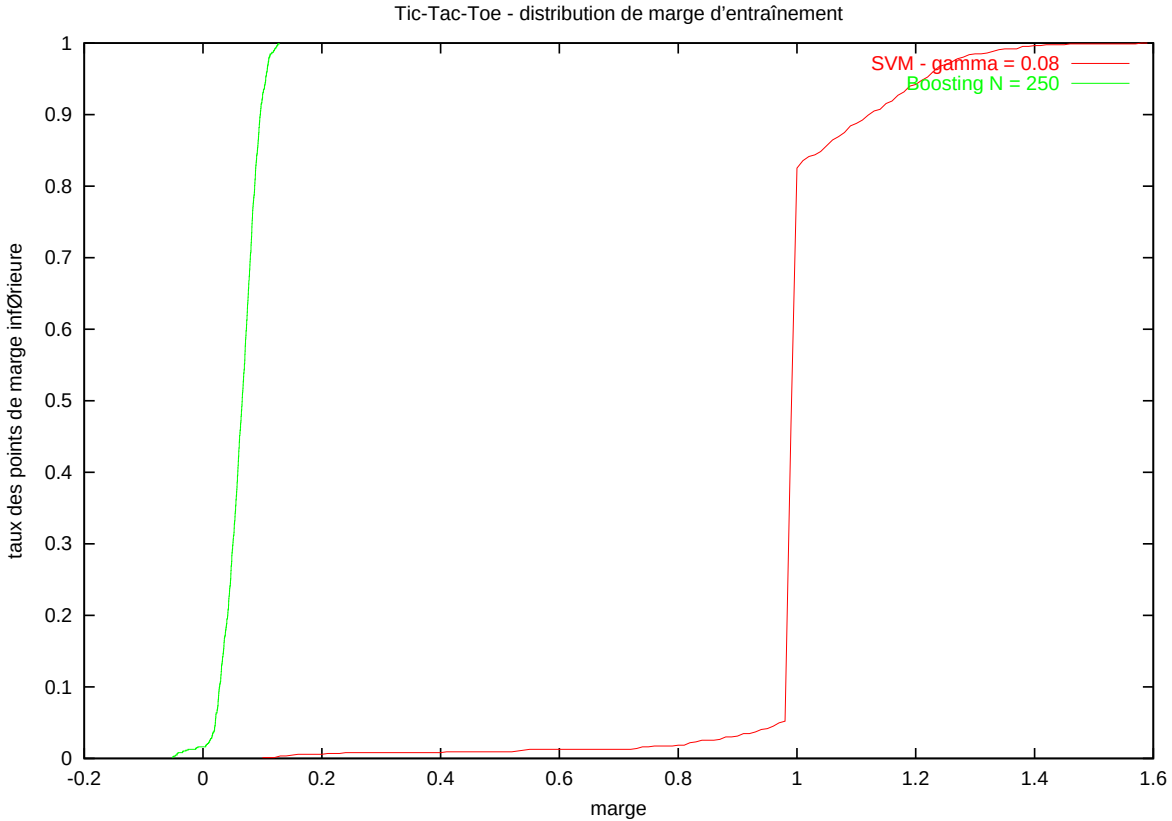
Machines à vecteurs de support

- Comparaison avec boosting (marge de test)



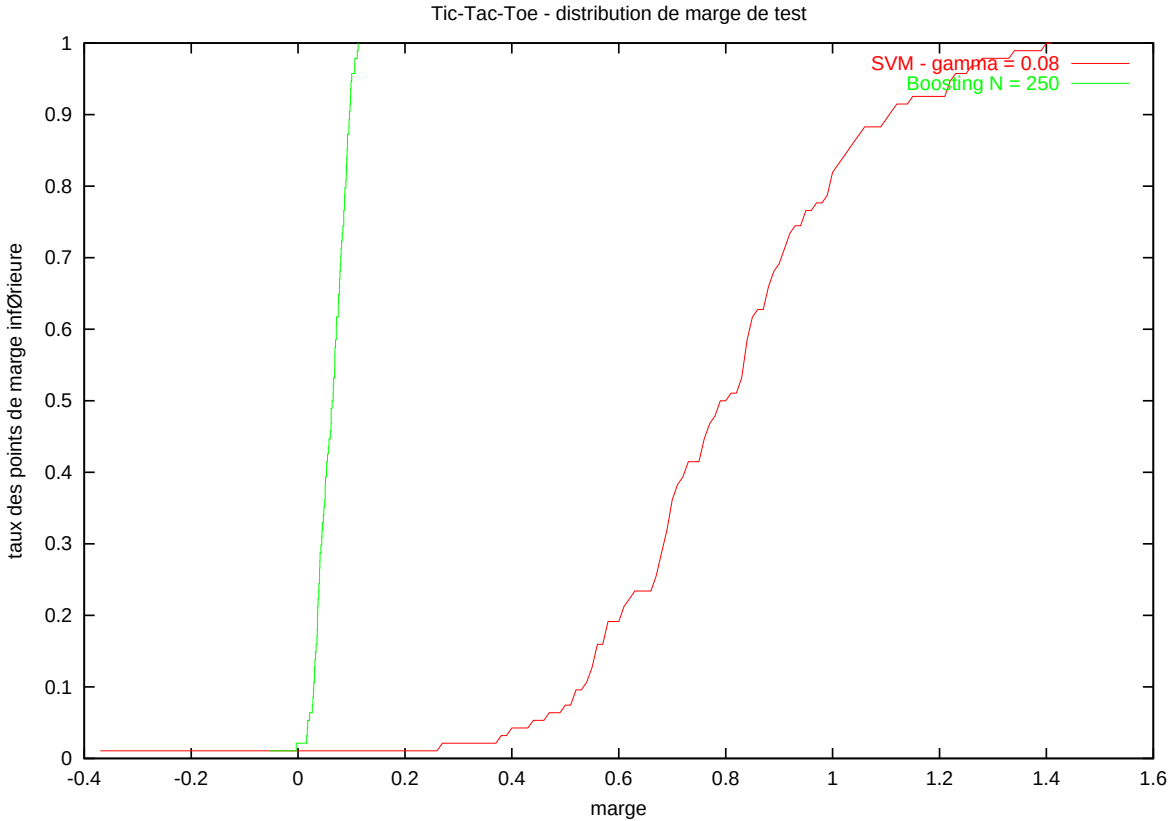
Machines à vecteurs de support

- Comparaison avec boosting (marge d'entraî nement)



Machines à vecteurs de support

- Comparaison avec boosting (marge de test)



Machines à vecteurs de support

- Erreur de généralisation
  - dimension VC effective:

$$\min\left(\frac{R^2}{\gamma^2}, d\right) + 1$$

  - espérance de l'erreur de généralisation:

$$\mathbb{E}[R(f_n)] \leq \frac{1}{n+1}\mathbb{E}\left[\frac{R_{n+1}^2}{\gamma_{n+1}^2}\right]$$

Machines à vecteurs de support

- Estimation LOO
  - erreur empirique:

$$\widehat{R}(f, D_n) = \widehat{R}(f) = \frac{1}{n}\sum_{i=1}^{n} I_{\{f(X_i)Y_i<0\}}$$

  - $f^{(i)}$ est entraîné sur

$$D^{(i)} = \big((X_1,Y_1),\ldots,(X_{i-1},Y_{i-1}),(X_{i+1},Y_{i+1}),\ldots,(X_n,Y_n)\big)$$

  - erreur LOO:

$$\widehat{R}_{LOO}(f) = \frac{1}{n}\sum_{i=1}^{n} I_{\{f^{(i)}(X_i)Y_i<0\}}$$

Machines à vecteurs de support

- Estimation LOO
  - $\widehat{R}_{LOO}(f)$ est presque non-biaisé:

$$\mathbb{E}\widehat{R}_{LOO}(f_{n+1}) = R(f_n)$$

  - erreur LOO:

$$\begin{aligned}\widehat{R}_{LOO}(f) &= \frac{1}{n}\sum_{i=1}^{n} I_{\{f^{(i)}(X_i)Y_i<0\}} = \frac{1}{n}\sum_{i=1}^{n} I_{\{-f^{(i)}(X_i)Y_i\geq 0\}} \\ &= \frac{1}{n}\sum_{i=1}^{n} I_{\{-f(X_i)Y_i+(f(X_i)-f^{(i)}(X_i))Y_i\geq 0\}} \\ &\leq \frac{1}{n}\sum_{i=1}^{n} I_{\{U^{(i)}-1\geq 0\}}\end{aligned}$$

  - où $U^{(i)} \geq (f(X_i) - f^{(i)}(X_i))Y_i$

Machines à vecteurs de support

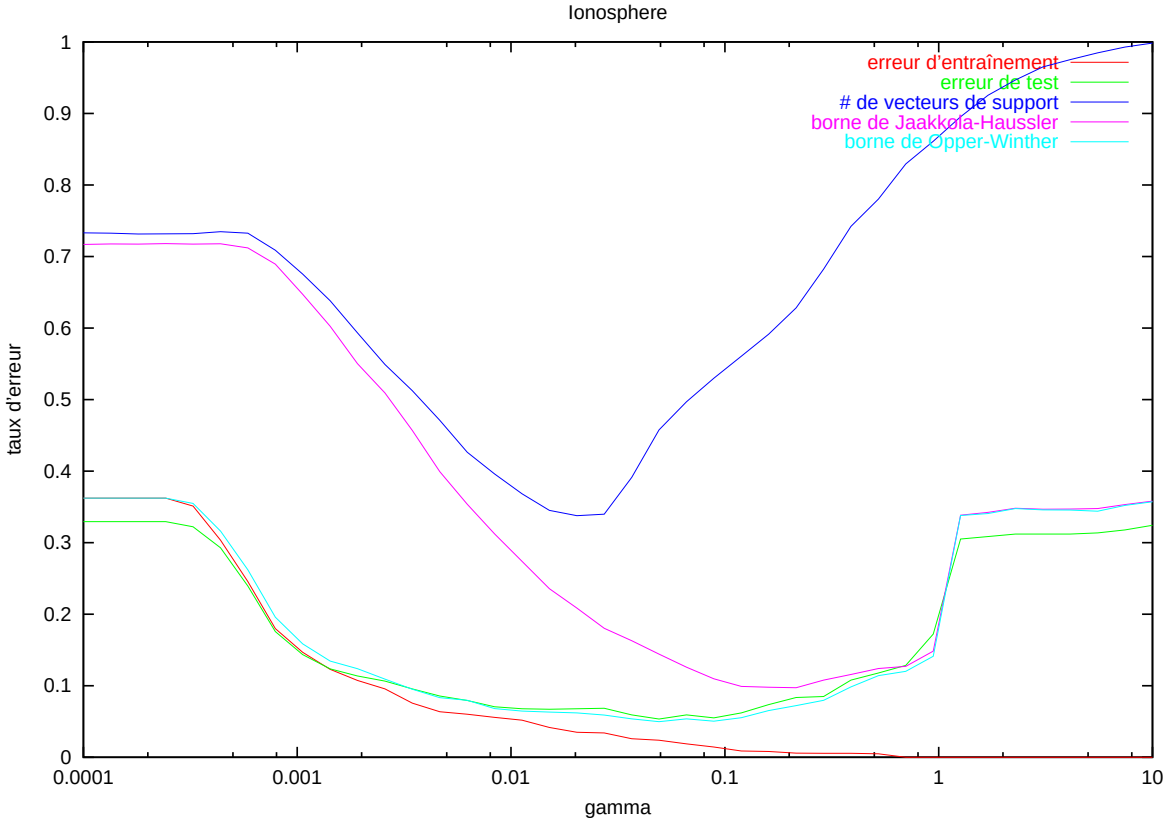
- Nombre de vecteurs de support
  - $U^{(i)} = 0$ pour les vecteurs de non-support
  - $\widehat{R}_{LOO}(f) \leq \frac{n_{SV}}{n}$

Machines à vecteurs de support

- Borne de Jaakkola-Haussler
  - $U^{(i)} \leq \alpha_i K(X_i, X_i)$ pour les vecteurs de non-support
  - $\widehat{R}_{LOO}(f) \leq \frac{1}{n} \sum_{i=1}^{n} I_{\{\alpha_i K(X_i, X_i) \geq 1\}}$
- Borne de Opper-Winther
  - matrice de Gram des vecteurs de support: $G_{SV\,ij} = K(X_i, X_j)$
  - $U^{(i)} = \frac{\alpha_i}{(G_{SV}^{-1})_{ii}}$
  - $\widehat{R}_{LOO}(f) \leq \frac{1}{n} \sum_{i=1}^{n} I_{\left\{ \frac{\alpha_i}{(G_{SV}^{-1})_{ii}} \geq 1 \right\}}$

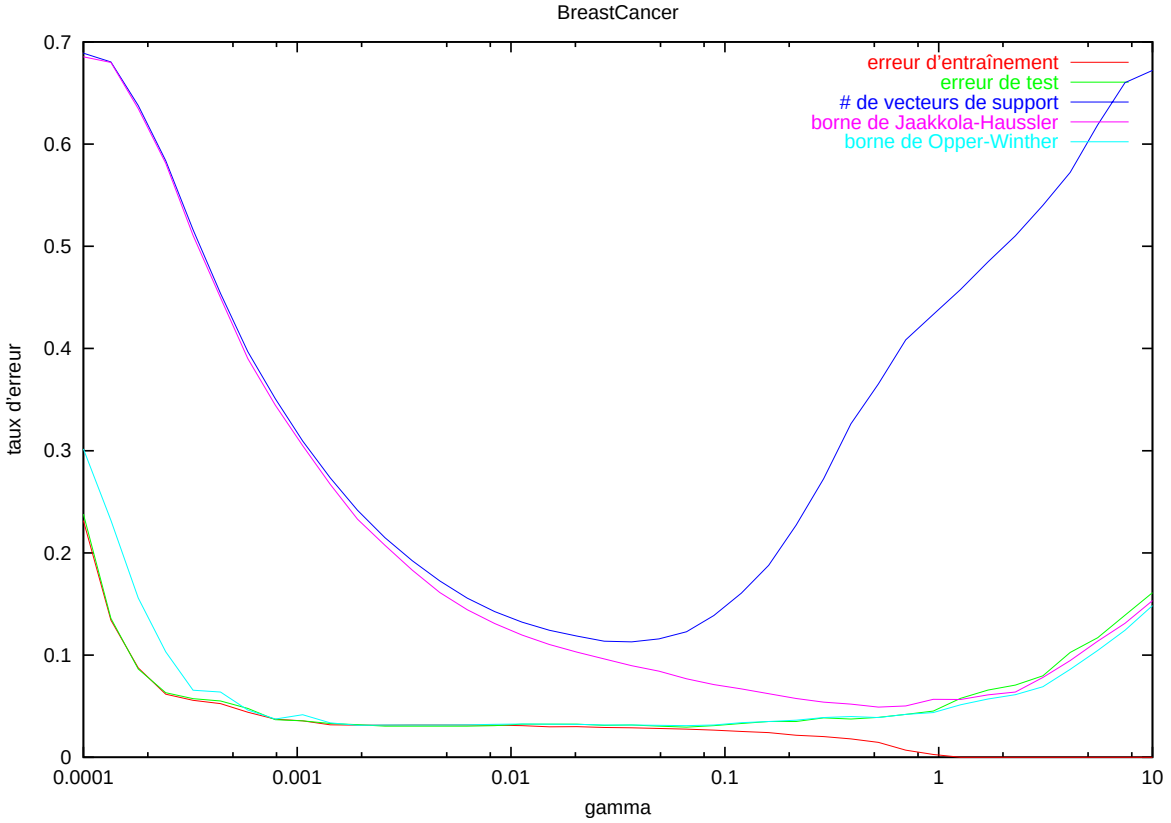
Machines à vecteurs de support

- Expériences



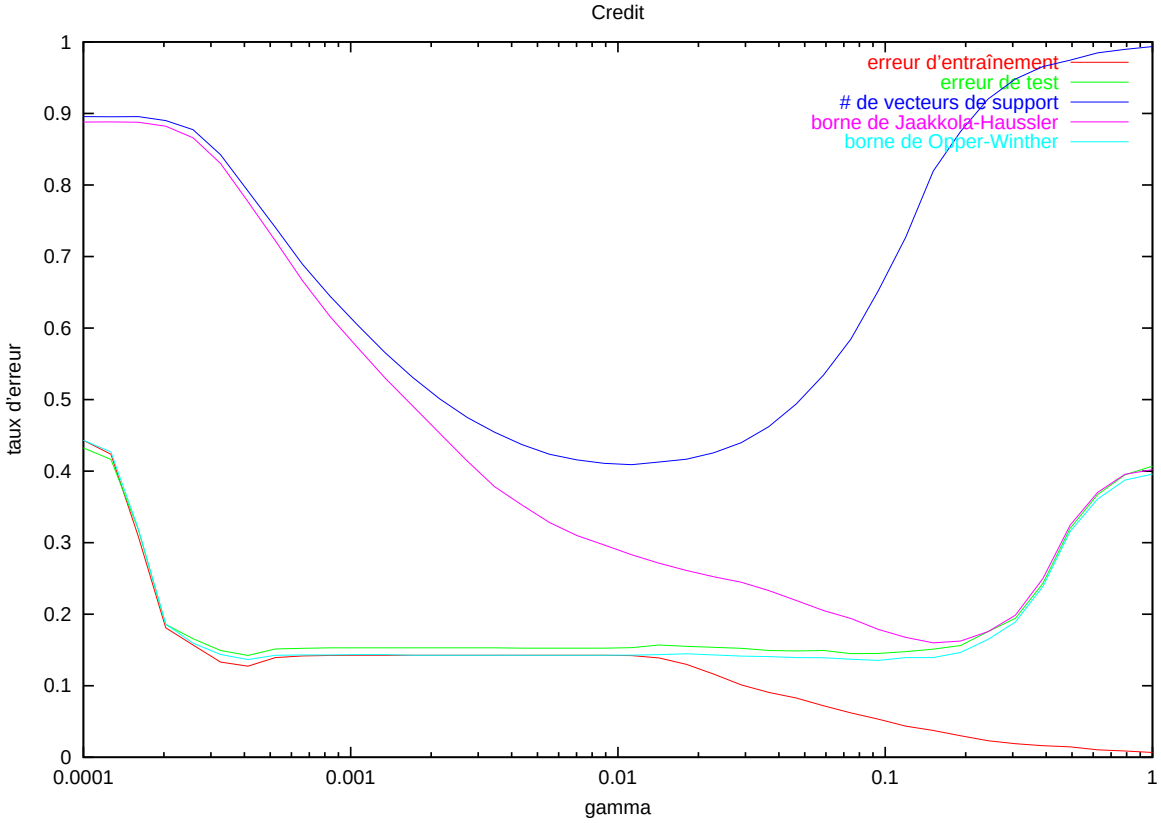
Machines à vecteurs de support

- Expériences



Machines à vecteurs de support

- Expériences



Machines à vecteurs de support

- Expériences