7. Sachant que :

$$(A^{-1}+vv^T)^{-1} = A - \frac{1}{(1+v^TAv)}Avv^TA \tag{8}$$

Montrer que :

$$Trace[(A^{-1}+vv^T)^{-1}] = Trace(A) - \frac{v^TAAv}{(1+v^TAv)} \tag{9}$$

8. Dans le modèle de régression linéaire simple que nous discutons la conséquence du rajout d'une donnée supplémentaire est-elle nécessairement positive ?

9. On revient au cas de la droite de regression avec des abscisses variant obligatoirement dans $[-1,1]$ et on suppose que X contient 2 abscisses d'entrée telles que $x_1 = 1$ et $x_2 = -1$. Que vaut X ? Quels sont (intuitivement) les meilleures valeurs de v (ou x) possibles lorsqu'on veut rajouter une 3e donnée selon le principe de l'apprentissage actif ?

10. Pour ce dernier X et $v = [x \;\; 1]^T$, on peut calculer A et montrer qu'en développant et en maximisant $\frac{v^TAAv}{(1+v^TAv)}$ (selon v ou x) on confirme l'intuition précédente.

2 Classifieur binaire optimal

On considère un classifieur dont l'espace d'entrée est $\mathcal{X} = [0.0, 1.0] = [0,1]$ (l'intervalle réel) et l'espace discret de sortie est $\mathcal{Y} = \{0,1\}$ (cas binaire à deux classes). On suppose de plus que la densité pour l'espace d'entrée est $P_X(x) = 1$, $\forall x \in \mathcal{X}$ et $Pr(Y = 1|X = x) = x$, $\forall x \in \mathcal{X}$. $\int dx$

1. Grâce à un dessin dans le plan $(\mathcal{X}, \mathcal{Y})$ et sans aucun calcul, on vous demande d'abord de donner deux allures vraisemblables pour *deux* ensembles d'apprentissage $D_1 = \{x_i, y_i\}_{i=1...6}$ et $D_2 = \{x'_i, y'_i\}_{i=1...6}$. Chaque ensemble comportera six exemples, il suggèrera bien la nature stochastique de la réalisation des échantillons *tout en respectant globalement les hypothèses sur les lois de probabilités* en entrée et en sortie. On vous demande de commenter vos propositions.

2. On vous demande de rappeler ce qu'est un prédicteur au plus proche voisin dans ce cas de classification binaire. Grâce à un dessin, on vous demande de donner l'allure de ce prédicteur $1ppv_{D_1}(x)$ bâti à partir de votre ensemble D_1.

3. Discuter (sans calcul) la capacité de généralisation de votre prédicteur $1ppv_{D_1}(x)$ relativement à l'ensemble D_2.

4. Trouver le classifieur optimal $Opt : \mathcal{X} \to \mathcal{Y}$ pour cette distribution et déterminer son risque $R(Opt) = EPE(Opt)$ lorsqu'on utilise la perte 0-1 e_z.

5. Pour un espace de sortie dénombrable $\mathcal{Y}$, la régression $r(x)$ est définie par

$$r(x) = E_{Y|X=x}[Y] = \sum_{y\in\mathcal{Y}} yPr(Y = y|X = x) \tag{10}$$

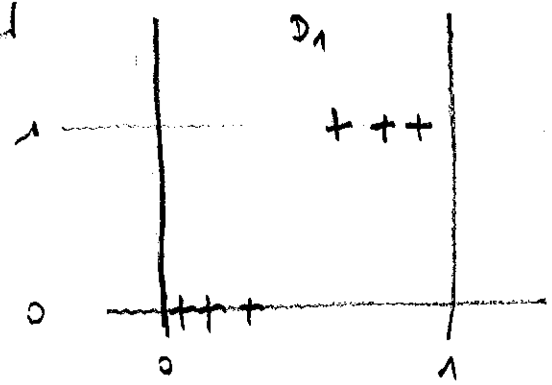
Que vaut $r(x)$ ici ? Donner une interprétation.

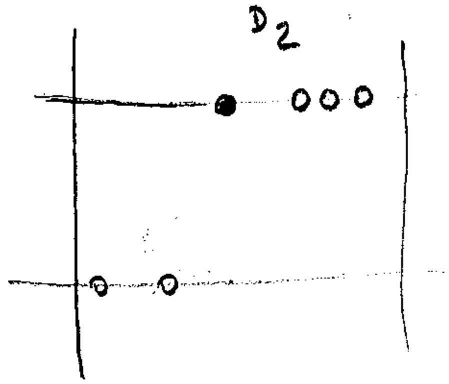
6. Calculer le risque quadratique de la régression que nous venons de calculer. Le risque est défini par :

$$R_r = \int_{\mathcal{X}} \sum_{y\in\mathcal{Y}} (r(x) - y)^2 Pr(Y = y|X = x)P_X(x)dx \tag{11}$$

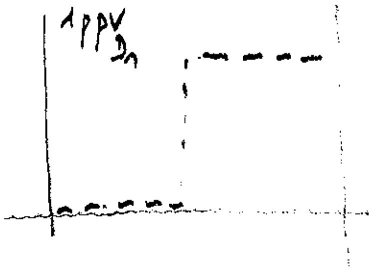
[ey02]

2.1





2.2



2.3 le point • de D_2 est mal prédit

2.4 $\mathrm{Opt}(x) = \underset{y \in \{0,1\}}{\operatorname{argmax}} \Pr(Y=y \mid X=x)$ pour la perte eg une informative

$$= \begin{cases} 1 & \text{si } \Pr[Y=1 \mid X=x] > \frac{1}{2} \\ 0 & \text{sinon} \end{cases} = \begin{cases} 1 & \text{si } x > \frac{1}{2} \\ 0 & \text{si } x \leq \frac{1}{2} \end{cases}$$

$$R(\mathrm{opt}) = \int_x \underset{p_X(x)}{1} . \left(e_g(\mathrm{opt}(x), 0)\, P[Y=0 \mid X=x] + e_g(\mathrm{opt}(x), 1)\, \underbrace{P[Y=1 \mid X=x]}_{x} \right) dx$$

$$= \int_0^{1/2} 1.1.x\, dx + \int_{1/2}^{1} 1.\,1.\,(1-x)\, dx = \left[\frac{x^2}{2}\right]_0^{1/2} + \left[x - \frac{x^2}{2}\right]_{1/2}^{1} = \frac{1}{4}$$

2.5 $r(x) = \sum_y y\, P[Y=y \mid X=x] = 1.\, P(Y=1 \mid X=x) = x$

$$R_r = \int_0^1 dx \sum_y P(Y=y \mid X=x)(x-y)^2$$

$$= \int_0^1 dx \left[P(Y=1 \mid X=x)(x-1)^2 + P(Y=0 \mid X=x)(x-0)^2 \right]$$

$$= \int_0^1 dx \left[x(x-1)^2 + (1-x)x^2 \right] = \int_0^1 dx\, (-x^2 + x) = \left[-\frac{x^3}{3} + \frac{x^2}{2} \right]_0^1 = \frac{1}{6}$$