Exercice 1 – Perceptron (examen 2009)

On s'intéresse à des classifieurs linéaires pour faire de la discrimination à deux classes. On note $D = \{x^i, y^i\}_{i=1,\dots,N}$ un ensemble d'apprentissage avec $y^i = 1$ si $x^i \in C_1$ et $y^i = -1$ si $x^i \in C_2$. $\mathbf{w}$ est le vecteur de poids du classifieur linéaire.

Q 1.1 Rappeler l'algorithme du perceptron

Q 1.2 On suppose que l'algorithme est initialisé avec $\mathbf{w}_0 = 0$

Q 1.2.1 Montrer qu'à une étape de l'algorithme, il existe des coefficients α_i tels que $\mathbf{w}_t = \sum_{i=1}^{N} \alpha_i y^i x^i$

Q 1.2.2 Exprimer en fonction des α_i la condition qui indique que le perceptron fait une erreur sur x_i.

Q 1.2.3 Reformuler l'algorithme du perceptron avec les seuls α_i comme paramètres.

Q 1.2.4 Exprimer la fonction de décision en fonction des α_i

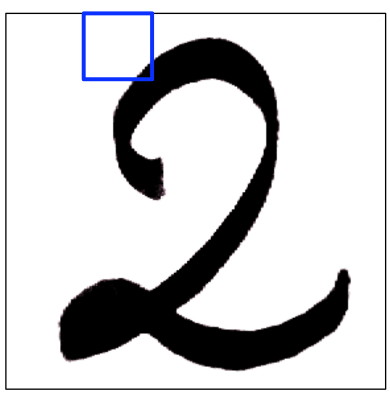
Q 1.3 On considère le critère d'apprentissage $Q(\mathbf{w}) = \sum_{i \in Z} -y^i x^i \mathbf{w}$ avec $Z = \{x^i | y^i x^i \mathbf{w} < 0\}$ l'ensemble des points mals classés par $\mathbf{w}$.

Q 1.3.1 Donner l'algorithme de gradient qui minimise ce critère.

Q 1.3.2 Donner l'algorithme de gradient adaptatif (=stochastique) qui minimise ce critère. Quel algorithme reconnaissez vous ?

Exercice 2 – Classification d'images binaires (exam 2010)

Soit une base d'images binaires (pixels noirs ou blancs) de taille 256x256 pixels. Un exemple d'image est fourni Fig. 2. Nous avons tenté au cours du semestre une analyse bayesienne pixel par pixel à l'aide d'une loi de Bernoulli. Malheureusement, les chiffres n'étaient pas tous positionnés de la même manière et cela à causer une perte de performances. Pour palier ce problème, nous proposons de recommencer l'expérience en utilisant une fenêtre glissante sur l'image. Une telle fenêtre est dessinée sur l'exemple.



La fenêtre peut prendre N positions dans l'image. Nous indicerons la position en i, x_i désigne la fenêtre dans la position i avec i variant de 1 à N. Pour une image donnée, la variable x_i prend la valeur k_i. k_i est le nombre de pixels noirs dans la fenêtre. x_i suit une loi Binomiale $\mathcal{B}(n, p_i)$. La fenêtre choisie est de taille 2x2 soit 4 pixels. Afin de représenter une image $\boldsymbol{x}$, nous utiliserons l'ensemble des fenêtres : $\{x_i\}_{i=1,\ldots,N}$.

1. Quelles valeurs peuvent prendre les x_i ?
2. Donner l'expression de la loi binomiale $p(x_i = k_i)$?. Que modélisent n et p_i ?
3. Soit $\mathbf{x}$ une image, $\mathbf{x} = \{x_1, \ldots, x_N\}$ donner l'expression de $p(\mathbf{x})$ si la valeur observée pour x_i est k_i.
4. On note $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_P\}$ un ensemble d'image $\mathbf{x}_j = \{x_{j,1}, \ldots, x_{j,N}\}$. Donner la log vraisemblance de X
5. Montrer que la vraisemblance est maximisée pour $p_i = \frac{\sum_{j=1}^{P} k_{j,i}}{nP}$
6. On veut résoudre un problème à deux classes avec cette approche, comment procède-t-on ?

Exercice 3 – Regression (exam 2011)

On considère un problème de régression linéaire. La cible est une variable t qui est une combinaison linéaire des variables d'entrée $\mathbf{x}$ plus un bruit blanc. Ce qui s'écrit : $t = y + \varepsilon$.
Avec :

$$y = \mathbf{w}^T\mathbf{x} = w_0x_0 + w_1x_1 + \ldots + w_nx_n$$

$$t \in \mathbb{R}, y \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^{n+1}, \varepsilon \sim \mathcal{N}(0, \beta^{-1})$$

$\beta = \frac{1}{\sigma^2}$, inverse de la variance est appelé *précision*. De manière générale, on note $\mathcal{N}(x|\mu, \sigma^2)$ la distribution gaussienne de moyenne μ et d'écart type σ d'une variable x. Dans toute la suite $x_0 = 1$, pour les vecteurs $\mathbf{x}$ (on a augmenté tous les vecteurs avec une première coordonnée 1). Soit $D = \{\mathbf{x}_i, t_i\}$ un ensemble d'apprentissage. On note $\mathbf{t} = (t_{1,\ldots,t_N})^T$ le vecteur colonne des sorties cibles et $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ l'ensemble des vecteurs d'entrée pour les données d'apprentissage.

Q 3.1 Préliminaires

Q 3.1.1 Montrer que pour des valeurs $\mathbf{x}, \mathbf{w}, \beta$ données, t suit une loi normale $\mathcal{N}(y, \beta^{-1})$. On notera par la suite $p(t|\mathbf{x}, \mathbf{w}, \beta) = \mathcal{N}(t|y, \beta^{-1})$ cette probabilité.
On remarque que t est la somme dune valeur constante y et d'une variable aléatoire gaussienne ε.

Q 3.1.2 Donner l'expression de $p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta)$

Q 3.1.3 Montrer que

$$\ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = \frac{N}{2}\ln\beta - \frac{N}{2}\ln(2\pi) - \beta E_D(\mathbf{w})$$

Avec :

$$E_D(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{N}(t_i - \mathbf{w}^T\mathbf{x}_i)^2$$

Q 3.2 Estimation du maximum de vraisemblance

Q 3.2.1 Calculer les gradients $\frac{\partial \ln p(\mathbf{t}|\mathbf{x},\mathbf{w},\beta)}{\partial \mathbf{w}}$ et $\frac{\partial \ln p(\mathbf{t}|\mathbf{x},\mathbf{w},\beta)}{\partial \beta}$

Q 3.2.2 Montrer que la solution à $\frac{\partial \ln p(\mathbf{t}|\mathbf{x},\mathbf{w},\beta)}{\partial \mathbf{w}} = 0$ est le vecteur $\mathbf{w}_{ML} = (XX^T)^{-1}X\mathbf{t}$

Q 3.2.3 ...