Devoir maison : Réseaux de neurones - classification

Bouacha Lazhar

Octobre 2020

Partie 1 :

Question 1 :

1.

$$m_p = \mathbb{E}[\| X - Y \|^2] = \sum_{i=1}^{p} \mathbb{E}[X_i^2] - 2\mathbb{E}[X_i]\mathbb{E}[Y_i] + \mathbb{E}[Y_i^2] \tag{1}$$

$$= \mathbb{E}[\| X \|^2] + \mathbb{E}[\| Y \|^2] = p(\mathbb{E}[X_1^2] + \mathbb{E}[Y_1^2]) = 2p \tag{2}$$

$$s_p^2 = Std[\| X - Y \|^2]^2 = \mathbb{V}(\| X - Y \|^2) = \mathbb{V}(\| Z \|^2), \text{ Z = X - Y} \tag{3}$$

$$= \sum_{i=1}^{p} \mathbb{V}(Z_i^2),\ Z_i \perp\!\!\!\perp Z_j, i \neq j \tag{4}$$

$$= p(\mathbb{V}(Z_1^2)) = p(\mathbb{V}(|X_1 - Y_1|^2)) = p(\mathbb{V}(|X_1^2 - 2X_1Y_1 + Y_1^2|)) \tag{5}$$

$$= p(\mathbb{V}(X_1^2) + 4\mathbb{V}(X_1)\mathbb{V}(Y_1) + \mathbb{V}(Y_1^2)) = p(2 + 4 + 2) = 8p \tag{6}$$

$$\mathbb{V}(X_1^2) = \mathbb{E}[X_1^4] - \mathbb{E}[X_1^2]^2 = 3 - 1 = 2 \tag{7}$$

$$\mathbb{E}[X_1^4] = \frac{(2.2)!}{2!2^2} = \frac{4!}{2!.4} = \frac{4 \times 3}{4} = 3, \text{ n = 2m = 4 donc m = 2} \tag{8}$$

$$\mathbb{E}[X_1^2] = 1^2 = 1 \tag{9}$$

$s_p = \sqrt{8p} = 2\sqrt{2p}$

2.

$$\frac{s_p}{m_p} = \frac{2\sqrt{2p}}{2p} = \frac{\sqrt{2p}}{p} \xrightarrow[p\to\infty]{} 0$$

3.

L'espérance croît linéairement en p alors que l'écart-type ne croît qu'en $\sqrt{p}$ donc si p est grand et qu'on tire n points i.i.d avec loi normale, tous les points ont tendances à être équidistants. La notion de plus proche voisin perdant son sens. Les points en grande dimension ont tendance à être isolés les uns des autres. Les méthodes d'apprentissage reposant sur la notion de plus proche voisin perdant donc une partie de leur sens en grande dimension. Exemple : k-NN

4.

On pourrait augmenter les observations, réduire l'espace ou régulariser.

Question 2 :

$$V_p(r) = \int_{B_p(r)} 1dx, y = \frac{x}{r} \Longrightarrow dy = \frac{dx}{r^p}, x \in B_p(r) \iff y \in B_p(1)$$

$$V_p(r) = \int_{B_p(1)} r^p dy = r^p \int_{B_p(1)} 1dy = r^p V_p(1) \tag{10}$$

$$V_1(1) = \frac{\sqrt{\pi}}{\frac{\sqrt{\pi}}{2}} = 2 = 2(1)(= 2.R) \tag{11}$$

$$V_2(1) = \frac{\pi}{\Gamma(2)} = \frac{\pi}{1} = \pi(= \pi.R^2) \tag{12}$$

2.a.

$\| X_i \| = d(0, X_i)$ et $x^p = e^{pln(x)}$, donc avec la fonction de répartition, F : $F(x) = \mathbb{P}(\| X_i \| \leq x) = \mathbb{P}(X_i \in B_p(x)) = \dfrac{Vol(B_p(x))}{Vol(B_p(1))} = 1$

2.b.

$$G(x) = \mathbb{P}(min(\| X_i \|) \leq x) = 1 - \mathbb{P}(min(\| X_i \|) > x) \tag{13}$$

$$= 1 - \mathbb{P}(\bigcap_{i=1}^{n} \{\| X_i \| > x\}) = 1 - \mathbb{P}(\| X_1 \| > x)^n \tag{14}$$

$$= 1 - (1 - x^p)^n \sim G(x^p), n \in \mathbb{N} \tag{15}$$

$x \in [0, 1] \Longrightarrow 1 - x^p \in [0, 1]$ et $min(\| X_i \|) \sim G(x^p) \Longrightarrow \| X_i \| \sim F(x) = x^p$

Donc en grande dimension avec p grand cette probabilité est proche de 0 et les points X_i de $B_p(1)$ se concentrent sur les bords de l'espace. C'est un problème lié à la grande dimension pour la prédiction car on doit extrapoler à partir des points au lieu d'interpoler pour prédire. Exemple : données météorologiques

Question 3 :

On a tiré aléatoirement θ puis on a concaténé un vecteur cosinus et sinus avant d'ajouter un vecteur aléatoire où X de taille 200x2.