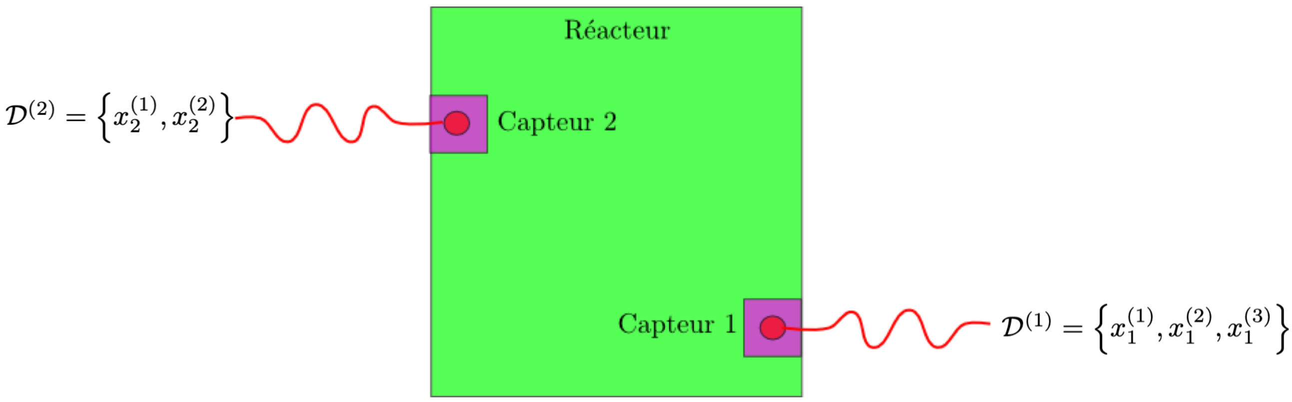
EnsLearn - TP n°2

Dans ce 2ème TP, on s'intéresse au problème connu sous le nom de **fusion de capteurs**. Par exemple, on souhaite connaître avec grande précision la température régnant dans un réacteur. On dispose $M = 2$ capteurs (potentiellement de technologies différentes) et, dans un intervalle de temps fini, ils délivrent chacun une serie de mesures. La taille des échantillons founis pas les deux capteurs n'est pas nécessairement la même.



Ce problème n'est pas à proprement parler un problème de comité de classifieurs/régresseurs, mais s'en rapproche fortement au sens où il s'agit d'estimer les paramètres des distributions ayant généré les données (nos relevés capteur). Cette démarche est très souvent celle employée en apprentissage où l'on va proposer une famille paramétrée pour $p_{X,Y}$ et chercher à estimer les paramètres de cette distribution grâce aux données.

Exercice 1 : Fusion de capteurs de variances connues

On suppose dans la suite du TP que les capteurs fournissent des mesures bruitées de la solution recherchée x. On suppose aussi que les distributions sont normales :

$$
\begin{aligned}
X_1 &\sim \mathcal{N}(x, \sigma_1) \quad (1)\\
X_2 &\sim \mathcal{N}(x, \sigma_2) \quad (2)
\end{aligned}
$$

où X_1 et X_2 sont deux variables aléatoires représentant les deux capteurs.

Dans cet exercice, on suppose les deux paramètres σ_1 et σ_2 connus. On note $\mathcal{D} = \left\{\mathcal{D}^{(1)}, \mathcal{D}^{(2)}\right\}$ l'ensemble des données. La vraisemblance s'écrit alors

$$
\begin{aligned}
\mathcal{L}(x) &= p(\mathcal{D}|x), \quad (3)\\
&= \prod_{k=1}^{n_1} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{\left(x_1^{(k)}-x\right)^2}{2\sigma_1^2}} \times \prod_{k=1}^{n_2} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{\left(x_2^{(k)}-x\right)^2}{2\sigma_2^2}} \quad (4)
\end{aligned}
$$

où n_1 est le cardinal de $\mathcal{D}^{(1)}$ et n_2 celui de $\mathcal{D}^{(2)}$.

Questions :

1. En prenant $x = 4$, $\sigma_1 = 0.1$ et $\sigma_2 = 0.3$, générez les propositions $\mathcal{D}^{(1)}$ et $\mathcal{D}^{(2)}$ de sorte que $n_1 = 4$ et $n_2 = 10$.

Une première méthode de fusion consiste à obtenir la valeur $\bar{x}$ qui maximise la vraisemblance globale et de trouver une expression faisant intervenir les maximiseurs des vraisemblances relatives à $\mathcal{D}^{(1)}$ et $\mathcal{D}^{(2)}$. Informatiquement, il est souvent plus commode de minimiser la *negative log-likelihood* (NLL) qui s'écrit

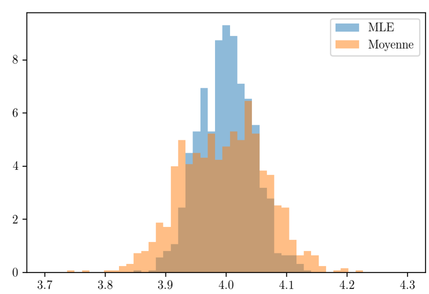
$$\begin{aligned} \mathrm{NLL}\,(x) &= -\log\, p\,(\mathcal{D}|x)\,, \qquad (5)\\ &= \sum_{k=1}^{n_1} \frac{\left(x_1^{(k)} - x\right)^2}{2\sigma_1^2} + \sum_{k=1}^{n_2} \frac{\left(x_2^{(k)} - x\right)^2}{2\sigma_2^2} + \mathrm{cte}. \qquad (6) \end{aligned}$$

Comme cette fonction est convexe, il s'en suit que sa dérivée s'annule en $\bar{x}$, d'où

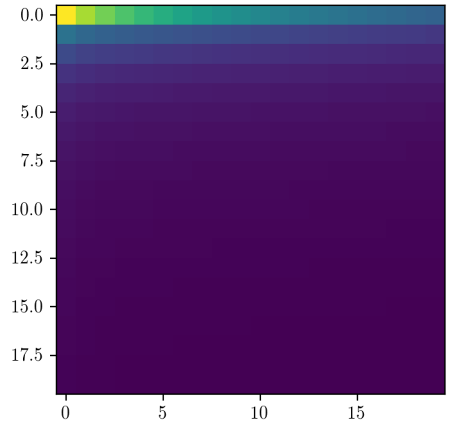
$$\begin{aligned} \sum_{k=1}^{n_1} -2\frac{x_1^{(k)} - \bar{x}}{2\sigma_1^2} + \sum_{k=1}^{n_2} -2\frac{x_2^{(k)} - \bar{x}}{2\sigma_2^2} &= 0, \qquad (7)\\ \Leftrightarrow \frac{1}{\sigma_1^2}\sum_{k=1}^{n_1}\left\{x_1^{(k)} - \bar{x}\right\} + \frac{1}{2\sigma_2^2}\sum_{k=1}^{n_2}\left\{x_2^{(k)} - \bar{x}\right\} &= 0, \qquad (8)\\ \Leftrightarrow \bar{x} &= \frac{\sigma_2^2\bar{x}_1 + \sigma_1^2\bar{x}_2}{\sigma_2^2 n_1 + \sigma_1^2 n_2}, \qquad (9) \end{aligned}$$

où $\bar{x}_1$ et $\bar{x}_2$ sont les moyennes empiriques respectives des 2 capteurs et sont les estimés du maximum de vraisemblance de $\mathcal{D}^{(1)}$ et $\mathcal{D}^{(2)}$ respectivement.

2. Pour 1000 réalisations, calculez l'agrégat $\bar{x}$ obtenu par maximum de vraisemblance et par simple moyenne arithmétique. Tracez les histogrammes (entre -0.3 et 0.3 pour 50 bins) relatifs aux deux méthodes. Discutez le résultat.
   Le résultat attendu est :



3. On généralise à présent l'expérience précédente à des cardinaux n_1 et n_2 variant de 1 à 20. Toujours pour 1000 réalisations, stockez dans un `2D numpy array` l'écart-type des agrégats $\bar{x}$ en fonction de n_1 et n_2. Affichez ensuite ce tableau à l'aide de `matplotlib.imshow` (avec l'option `interpolation='none'`).
   Le résultat attendu est :



Discutez de l'attrait de la capteur n°1 par rapport au n°2.

On suppose à présent avoir des connaissances *a priori* sur la température du réacteur avant même d'avoir vu des données. Ces connaissances sont exprimées sous forme d'une densité de probabilité normale elle aussi :

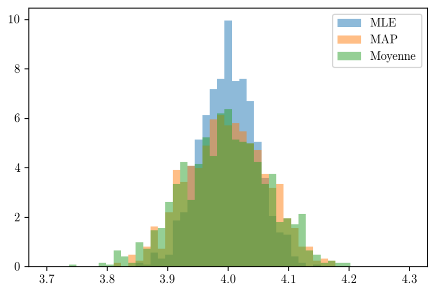
$$p\left(x\right) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(x_0 - x)^2}{2\sigma_0^2}}. \tag{10}$$

L'agrégat s'exprime alors comme le maximum *a posterior* (MAP), c'est à dire la valeur maximisant $p\left(\mathcal{D}|x\right) \times p\left(x\right)$. On montre que dans ce cas,

$$\bar{x} = \frac{\frac{\bar{x}_1}{\sigma_1^2} + \frac{\bar{x}_2}{\sigma_2^2} + \frac{x_0}{\sigma_0^2}}{\frac{n_1}{\sigma_1^2} + \frac{n_2}{\sigma_2^2} + \frac{1}{\sigma_0^2}} \tag{11}$$

3. On suppose que le prior $p\left(x\right)$ provient d'une étape précédente de calcul, où nous avions $n_1 = n_2 = 3$. Dans ces circonstances, x_0 est l'estimation obtenu avec ces données là par la méthode du maximum de vraisemblance. Le paramètre σ_0 est une entrée du `2D numpy array` calculé à la question précédente.

   On suppose également recevoir de nouvelles données dans un second temps toujours dans les mêmes proportions $n_1 = n_2 = 3$. Mettez à jour nos croyances sur x selon le principe du MAP. Sur 1000 réalisations, comparez à l'approche *batch* où nous aurions obtenu les 12 données d'un coup. Comparez également à la moyenne.
   Le résultat attendu est :



Discutez ce résultat.

Exercice 2 : Fusion de capteurs de variances inconnues

Dans ce second exercice, on considère à présent la situation où les variances des capteurs sont inconnues. La fonction de vraisemblance devient alors une fonction des trois paramètres x, σ_1 et σ_2. Son expression reste néanmoins inchangée et la propriété de convexité reste valable.
Pour obtenir l'agrégat, il faut tout de même calculer le maximum de vraisemblance pour les deux écarts-type. Sans surprise, les estimés MLE de ces deux paramètres sont les écarts-type empiriques capteur par capteur :

$$\sigma^2_{1,mle} = \frac{1}{n_1}\sum_{k=1}^{n_1}\left(x_1^{(k)} - \mathcal{D}\right)^2, \tag{12}$$

$$\sigma^2_{2,mle} = \frac{1}{n_2}\sum_{k=1}^{n_2}\left(x_2^{(k)} - \mathcal{D}\right)^2. \tag{13}$$

Ces deux expressions font intervenir l'agrégat $\mathcal{D}$ qui à présent se calcule en faisant intervenir lui aussi ces deux estimés :

$$\bar{x} = \frac{\sigma^2_{2,mle}\bar{x}^{(1)} + \sigma^2_{1,mle}\bar{x}^{(2)}}{\sigma^2_{2,mle}n_1 + \sigma^2_{1,mle}n_2} \tag{14}$$

On peut alors montrer que pour un triplet donné $(\bar{x}, \sigma_{1,mle}, \sigma_{2,mle})$, ce système d'équation équivaut à l'application d'une contraction dont le **point fixe** est la solution au problème.

1. Reprenez le même problème qu'à la question 2 de l'exercice précédent mais en estimant conjointement $(\bar{x}, \sigma_{1,mle}, \sigma_{2,mle})$ selon le mécanisme évoqué ci-dessus.