Exam solutions: Statistical modeling and its applications

Mines St-Etienne – 21st January 2015

Exercise 1 (5 pts)

On considère la fonction $f(x_1, x_2) = e^{x_1} x_2$. On souhaite réaliser une analyse de sensibilité globale de $f(X_1, X_2)$ lorsque X_1 et X_2 sont des variables aléatoires indépendantes uniformes sur [-1/2, 1/2].

1. Par symétrie de la densité de la loi unforme sur [-1/2, 1/2] on a $E(X_2) = 0$.

2. Il y a deux méthodes possibles ici, soit calcul direct, soit proposer une décomposition et vérifier qu'elle satisfait les hypothèses d'indépendances et de non-simplification (voir TDs). On obtient $\mu_0 = 0$, $\mu_1(x_1) = 0$, $\mu_2(x_2) = m_1 x_2$, $\mu_{1,2} = (e^{x_1} - m_1)x_2$

3. -

   (a) En utilisant l'indépendance de X_1 et X_2, on a $\text{var}(f(X_1, X_2)) = E((e^{X1})^2)E(X_2^2) = (m_1^2 + v_1)v_2$

   (b) $S_1 = 0$ est évident. $D_2 = \text{var}(\mu_2(X_2)) = m_1^2 v_2$ d'où l'expression attendue.

   (c) La somme des indices de Sobol vaut 1 donc $S_{1,2} = 1 - S_2 = \frac{v_1}{{m_1}^2 + v_1}$.

Exercise 2 (3 pts)

1. X_2, X_3, X_5 sont sont influentes (hiérarchie : X_2, X_5, X_3) ; les autres n'inter-viennent pas du tout car leur effet total est nul. X_2 interagit avec les variables influentes, donc seulement avec X_3 et X_5 (le graphique ne permettant pas d'aller plus loin). Les variables individuelles expliquent 80% de la variance, le reste étant des interactions.

2. Question de cours. Il faut dans l'ordre évaluer f aux simulations, ce qui donne un vecteur y de taille N. Sa moyenne donne une estimation de μ_0. On représente ensuite $y - \mu_0$ contre x_2 et on effectue un lissage, ce qui donne une estimation (moyenne mobile, splines, polynomes locaux) de $E(f(X_1, ..., X_8)|X_2) - \mu_0$.

Exercise 3 (5 pts)

1. Given the smoothness of the mean predictor, which is a linear combination of $k(x, X)$, the kernel could be squared exponential or Matern 5/2.

2. The smallest lengthcales ($\theta = 0.05$) can definitely be observed in the direction x_1 for the model m_1 since the model comes back very quickly to a mean value. Furthermore, the model m_2 seems to be isotropic, which is consistent with $(\theta_1, \theta_2) = (0.25, 0.25)$ for this model.

3. Let μ and C be the predicted mean values and covariance matrix for the test-set according to the model:

$$\mu = k(X_{test}, X)k(X, X)^{-1}F$$
$$C = k(X_{test}, X_{test})k(X_{test}, X)k(X, X)^{-1}k(X, X_{test}).$$

   The standardised residuals R are given by: $R = C^{-1/2}(F_{test} - \mu)$, where $C^{-1/2}$ is given, for example by the inverse of the Cholesky decomposition of C.

4. The second model should be preffered to the first one for two reasons :
    - Its standardised residuals are actually close to a standart Gaussian distribution
    - It has a better Q_2.
5. The first model comes back to a constant value (not equal to zero) when it is far from the observations, so there is a constant mean. Since we don't know the actual value of the variance parameter, it is not possible to say if the confidence intervals include an extra term accounting for the variance of the trend estimate. We thus don't know if it has been estimated or not.

Exercise 4 (2 pts)

- Z_1 has a Matern 3/2 kernel with variance $\sigma^2 = 1$ and lengthcale $\theta = 0.2$
- Z_2 has a composite kernel which is a sum of a Gaussian kernel $((\sigma^2, \theta) = (1.2)$ and a white noise kernel with a small variance ($\sigma^2 = 0.01$).
- Z_3 has a Brownian motion kernel with variance $\sigma^2 = 1$.
- Z_4 has a kernel of the shape constant plus linear : $\sigma_0^2 + \sigma_1^2 xy$, with $\sigma_0^2 = 0.05$ and $\sigma_1^2 = 1$.

Exercise 5 (3 pts)

1. Since $(x, y) \mapsto 1$ is a valid covariance function, k_{anova} can be seen as the sums and product of covariance functions. It is thus also a valid kernel.
2. Expanding the kernel expression gives

$$k_{anova}(x, y) = \sigma^2 \left(1 + \sum_{i=1}^{d} k_i(x_i, y_i) + \sum_{i=1}^{d} \sum_{i=1}^{d} k_i(x_i, y_i) k_j(x_j, y_j) + \cdots + \prod_{i=1}^{d} k_i(x_i, y_i) \right)$$

If we plug this in the best predictor expression, we obtain

$$\begin{aligned} m(x) &= k_{anova}(x, X) k_{anova}(X, X)^{-1} F \\ &= \sigma^2 \left(1 + \sum_{i=1}^{d} k_i(x_i, X_i) + \cdots + \prod_{i=1}^{d} k_i(x_i, X_i) \right) k_{anova}(X, X)^{-1} F \\ &= m_0 + \sum_{i=1}^{d} m_i(x_i) + \sum_{i=1}^{d} \sum_{i=1}^{d} m_{i,j}(x_i, x_j) \cdots + m_{1 \ldots d}(x) \end{aligned}$$

where $m_0 = 1 k_{anova}(X, X)^{-1} F$, $m_i(x_i) = k_i(x_i, X_i) k_{anova}(X, X)^{-1} F$, etc. The decomposition of the kernels shows that Z can be seen as the sum of independent GPs:

$$Z(x) = Z_0 + \sum_{i=1}^{d} Z_i(x_i) + \sum_{i=1}^{d} \sum_{i=1}^{d} Z_{i,j}(x_i, x_j) \cdots + Z_{1 \ldots d}(x)$$

Given this decomposition, the submodels can be interpreted as $m_I(x_I) = \mathrm{E}[Z(x_I) | Z(X) = Z]$ (I can be any subset of index).
3. The difference between the above decomposition of m in submodels and its Sobol-Hoeffding decompostion is that the m_I do not necessarily integrate to zero. This can be acheved by choosing the univariate kernels k_i as kernels generating sample paths that intergate to zero, as it has been discussed during the course.

Exercise 6 (2 pts)

Let $\mathcal{H}$ be a RKHS of functions over $[0,1]$ such that the derivative evaluation in $0 :\ L : f \mapsto f'(0)$ is continuous.

1. Since it is already stated in the wording that L is continuous, we just need to stress that computing a derivative is a linear operator in order apply the Riesz theorem:

$$L(f+\lambda g) = \frac{\mathrm{d}}{\mathrm{d}x}(f+\lambda g)\bigg|_0 = \frac{\mathrm{d}f}{\mathrm{d}x}\bigg|_0 + \lambda\frac{\mathrm{d}g}{\mathrm{d}x}\bigg|_0 = L(f) + \lambda L(g)$$

   Let r be its Riesz representer: Since $r \in \mathcal{H}$, we have $r(x) = \langle r, k(x,.)\rangle_{\mathcal{H}} = L(k(x,.)) = \frac{\mathrm{d}k(x,s)}{\mathrm{d}s}(0)$.

2. The subspace $\mathcal{H}_0$ corresponds to the subspace orthogonal to r so we have $k_0 = k - k_r$ where k_r is the reproducing kernel of $Span(r)$. Furthermore, the expression k_r is:

$$k_r(x,y) = \frac{r(x)r(y)}{r'(0)}$$