Appendix to Random projections in model selection and related experiment design problems

Ewa Skubalska-Rafajłowicz and Ewaryst Rafajłowicz

December 16, 2012

1 Appendix

Proof of Corollary 2. $\mathbf{M}_{\mathbf{v}}(\tilde{\xi}) = \mathbf{I}_r$ and $\sup_{\mathbf{x}\in\mathbf{X}} \mathbf{v}^T(\mathbf{x})\, M_{\mathbf{v}}^{-1}(\xi^*)\, \mathbf{v}(\mathbf{x}) = r$ is attained at points of the form $[\pm 1,\, \pm 1, \ldots,\, \pm 1]^T$ for which simultaneously both $||\mathbf{w}_L(\mathbf{x})||^2$ and $||\mathbf{w}_R(\mathbf{x})||^2$ attain their maxima that are equal to $\tilde{K}_L$ and $\tilde{K}_R$, respectively. At the same time, if κ is the support point, then $\mathbf{w}_L(\kappa) = \tilde{K}_L$ and $W_L(\tilde{\xi}) = \tilde{K}_L$. Analogously, $W_R(\tilde{\xi}) = \tilde{K}_R$. Thus, the maxima of $\phi(\mathbf{x},\, \tilde{\xi}) = r+2$, which finishes the proof. $\diamond$

Proof of Corollary 3. From the proof of Corollary 2 we known that $\mathbf{M}(\tilde{\xi})$ is $(r+2)\times(r+2)$ diagonal matrix with r diagonal entries equal 1 and the last two equal to $\tilde{K}_L$ and $\tilde{K}_R$, respectively. On the other hand, $\mathbf{M}(\xi_u)$ is also the diagonal matrix. Indeed, the off diagonal entries of $\mathbf{M}_{\mathbf{v}}(\xi_u)$ has the form:

$$2^{-d}\int_{-1}^{1}\ldots\int_{-1}^{1} x^{(k)} \ldots \int_{-1}^{1} x^{(l)} \ldots \int_{-1}^{1} dx^{(1)}\, dx^{(2)} \ldots dx^{(d)} = 0 \tag{1}$$

for $k\neq l$, with obvious change of $x^{(k)}$ to 1 when $x^{(l)}$ meets the constant term. For the diagonal entries of this matrix we have:

$$2^{-d}\int_{-1}^{1}\ldots\int_{-1}^{1} \left[x^{(k)}\right]^2 \ldots \int_{-1}^{1} dx^{(1)} \ldots dx^{(d)} = 2^{-d}\,\frac{2^d}{3} = \frac{1}{3} \tag{2}$$

with the obvious exception of the first diagonal entry, for which we replace $\left[x^{(k)}\right]^2$ by 1 in (2), which gives 1 as its value.

Finally, for the last two diagonal entries we have, provided that only the second order interaction terms are present in $\mathbf{w}(\mathbf{x})$,

$$W_L(\xi_u) = 2^{-d}\sum_{k=1}^{K_L}\int_{\mathbf{X}} \left[w^{(k)}(\mathbf{x})\right]^2\, d\mathbf{x}, \tag{3}$$

where each $\int_{\mathbf{X}} \left[w^{(k)}(\mathbf{x})\right]^2 d\mathbf{x}$ term has the following form:

$$\int_{-1}^{1}\ldots\int_{-1}^{1}\left[x^{(k_1)}\right]^2 \ldots \int_{-1}^{1}\left[x^{(k_4)}\right]^2 \ldots \int_{-1}^{1} dx^{(1)} \ldots dx^{(d)} = \frac{2^d}{3^4}. \tag{4}$$

Thus, $W_L(\xi_u) = \tilde{K_L}/3^4$. Analogously, for $W_R(\xi_u)$ we obtain $\tilde{K_R}/3^4$.

Summarizing, $\mathbf{M}(\xi_{mix}) = (1-\gamma)\,\mathbf{M}(\tilde{\xi}) + \gamma\,\mathbf{M}(\xi_u)$ is $(r+2)\times(r+2)$ diagonal matrix with the first diagonal entry 1, the next $d = r-1$ entries of the form: $(1-\gamma)+\gamma/3 = 1-2\,\gamma/3$ and the last two entries equal to $(1-\gamma)\,\tilde{K}_L + \gamma\,\tilde{K}_L/3^4$ and $(1-\gamma)\,\tilde{K}_R + \gamma\,\tilde{K}_R/3^4$, respectively. Hence, for the determinant of $\mathbf{M}(\xi_{mix})$ we have:

$$|\mathbf{M}(\xi_{mix})| = \tilde{K}_L\,\tilde{K}_R\,(1-\gamma\,80/81)^2\,(1-2\gamma/3)^d, \tag{5}$$

while, the determinant of $|\mathbf{M}(\tilde{\xi})| = \tilde{K}_L\,\tilde{K}_R$. Thus, D-efficiency $D_{eff}(d,\,\gamma)$ of ξ_{mix} is given by (**??**). $\diamond$

Description of the method of using random projections for selecting terms.

Step 0 Collect observations $(\mathbf{x}_i,\, y_i)$, i=1, 2,..., n. Select vectors $\mathbf{v}(\mathbf{x})$ and $\mathbf{w}(\mathbf{x})$. Select parameters: $\sigma_s > 0$ ($\sigma_s = 3$), working significance level $0 < \alpha < 1$ ($\alpha = 0.1$), final check significance level $0 < \alpha_f < 1$ ($\alpha_f = 0.05$), the number of random projections $rep \geq 1$ ($rep = 200$), the threshold $0 < \theta < 1$ ($\theta = 0.2$) as the fraction of (local) trials for stopping the algorithm. Set counter $c_0 = 0$. Prepare three empty lists *candidates* (the of the terms to be added to a regression function), *prospective* (the of the terms worth to be considered as the most perspective) and *waiting* (the of the terms to be considered later).

Step 1 Draw at random $\mathbf{s} \in \mathbf{R}^{\widetilde{K}}$, $\mathbf{s} \sim \mathcal{N}(\mathbf{0},\, \sigma_s^2\,\mathbf{I}_{\widetilde{K}})$. Form regression function (**??**) and find estimates $\hat{\mathbf{a}}_{\mathbf{s}}$ and $\hat{\beta}(\mathbf{s})$ by OLS. Test the hypothesis: $\hat{\beta}(\mathbf{s}) = 0$ at the level α. If the hypothesis is rejected $c_0 = c_0 + 1$.

Step 2 Repeat Step 1 *rep* times. If $c_0 < \theta\, rep$, STOP with the message: *probably there are no terms from* $\mathbf{w}(\mathbf{x})$ *to be added*, otherwise, go to Step 3.

Step 3 Enter all the terms from $\mathbf{w}(\mathbf{x})$ to *prospective* list.

Step 4 Split *prospective* list in half. Replace $\mathbf{w}_L(\mathbf{x})$ in (**??**) by the left part of this list and $\mathbf{w}_R(\mathbf{x})$ by the right half. Set counters $c_L = 0$, $c_R = 0$.

Step 5 Generate random Gaussian vectors $\mathbf{s}_L$ and $\mathbf{s}_R$ of the same lengths as the current $\mathbf{w}_L(\mathbf{x})$ and $\mathbf{w}_R(\mathbf{x})$ and find the estimates $\hat{\beta}_L(\mathbf{s}_L)$ and $\hat{\beta}_R(\mathbf{s}_R)$ by minimizing (**??**). Test the hypothesis: $\hat{\beta}_L(\mathbf{s}_L) = 0$ (respectively, $\hat{\beta}_R(\mathbf{s}_R) = 0$) and set $c_L = c_L + 1$ (respectively, set $c_L = c_L + 1$), if it is rejected.

Step 5 Repeat Step 5 *rep* times. If $c_L < \theta\, rep$ AND $c_R < \theta\, rep$, go to Step 13. Otherwise, if $c_L \geq c_R$ and

Step 5a if current $\mathbf{w}_L(\mathbf{x})$ contains more than one term, then replace all the content of *prospective* list by $\mathbf{w}_L(\mathbf{x})$ and add $\mathbf{w}_R(\mathbf{x})$ to *waiting* list, but only if $c_R \geq \theta\, rep$, otherwise, reject $\mathbf{w}_R(\mathbf{x})$ from considerations and go to Step 4,

Step 5b if current $\mathbf{w}_L(\mathbf{x})$ contains exactly one term, than add it to *candidate* list and add $\mathbf{w}_R(\mathbf{x})$ to *waiting* list, but only if $c_R \geq \theta\, rep$, otherwise, reject $\mathbf{w}_R(\mathbf{x})$ from further considerations. Then replace the content of the *prospective* list by all the *waiting* list, set *waiting* list to be empty and go to Step 4.

If $c_L < c_R$, perform Steps 5a) and 5b), replacing the roles $\mathbf{w}_L(\mathbf{x})$ and $\mathbf{w}_R(\mathbf{x})$.

Step 6 Final cleaning: if list *candidates* is empty, STOP with the message: *probably there are no terms from* $\mathbf{w}(\mathbf{x})$ *to be added*, otherwise, add this list to $\mathbf{v}(\mathbf{x})$, estimate the parameters of the extended regression function and test their significance at the level α_f. Reject nonsignificant terms, re-estimate parameters and provide the final list of regression terms.