Preliminary Analysis of MKL
Flat Maxima, Diversity and Fisher Information

Theo Damoulas
Institute for Computational Sustainability
Computing and Information Science
Cornell University
Ithaca, NY, USA

Mark A. Girolami
Inference Research Group
Dept. of Computing Science & Dept. of Statistics
University of Glasgow
Scotland, UK

Simon Rogers
Inference Research Group
Department of Computing Science
University of Glasgow
Scotland, UK

Understanding Multiple Kernel Learning Methods
NIPS Workshop, Whistler, December 2009
Preliminary Analysis of MKL

Flat Maxima, Diversity and Fisher Information

Flat Maximum Effect

Ambiguity Decomposition

Bias-Variance-Covariance

Design Of Experiments

Fisher Information
Flat Maximum Effect


\[ w = \sum_{i=1}^{d} w_i x_i \quad u = \sum_{i=1}^{d} u_i x_i \quad w_i, u_i \geq 0 \quad \sum_{i=1}^{d} w_i = \sum_{i=1}^{d} u_i = 1 \]

\[ \rho(u, w) \geq \sum_{i,j=1}^{d} u_i w_j \rho(x_i, x_j) \]

“Often quite large deviations from the optimal set of weights will yield predictive performance not substantially worse than the optimal weights”

\[ y_{e(b)} = \sum_{j=1}^{N} w_j \sum_{s=1}^{S} b_s k_s(x_i, x_j) = \sum_{s=1}^{S} b_s y_s \quad y_{e(\beta)} = \sum_{j=1}^{N} w_j \sum_{\sigma=1}^{S} \beta_\sigma k_\sigma(x_i, x_j) = \sum_{\sigma=1}^{S} \beta_\sigma y_\sigma \]

Flat Maximum Lower Bound

\[ \rho(y_{e(b)}, y_{e(\beta)}) \geq \sum_{s,\sigma=1}^{S} b_s \beta_\sigma \rho(y_s, y_\sigma) \]

The correlation between any two weighted kernel combination responses is lower bounded by a function of the correlation between base kernel responses.
- Ensemble response correlation increases as kernel alignment increases.
- Little or no benefit in parameterized [global] kernel combinations for highly aligned base kernels.
Decompositions of the Loss

\[
(y_e - y)^T (y_e - y) = \sum_{s=1}^{S} \beta_s (y_s - y)^T (y_s - y) - \sum_{s=1}^{S} \beta_s (y_s - y_e)^T (y_s - y_e)
\]

• Need for accurate (Weighted Ind. Error) but diverse (Ambiguity) base kernels.

\[
E \left\{ (y_e - y)^T ((y_e - y)) \right\} = \sum_{s, \sigma=1}^{S} \beta_s \beta_{\sigma} (E \{ y_s \} - y)^T (E \{ y_{\sigma} \} - y)
\]

\[
+ \sum_{s=1}^{S} \beta_s^2 \left( (y_s - E\{y_s\})^T (y_s - E\{y_s\}) \right) + \sum_{s=1, \sigma \neq s}^{S} \beta_s \beta_{\sigma} \left( (y_s - E\{y_s\})^T (y_{\sigma} - E\{y_{\sigma}\}) \right)
\]

• Fitting (Bias) versus Generalisation (Variance) versus Diversity (Covariance).
• Sample data from ensemble response.

• Quantifying the need for **accurate** and **diverse** base kernels.
• Sampling data from single base kernel response.

• Varying composite error - high individual error - diverse but not accurate.
Design of Experiments


- Maximize the information offered for the model parameters with respect to the evidence observed

\[
F(\theta) = -\mathbb{E} \left\{ \frac{\partial^2 \mathcal{L}}{\partial \theta \partial \theta^T} \right\}
\]

\[
\mathcal{L} = \left( y^T \sum_{s=1}^{S} \beta_s K_s w - \frac{1}{2} w^T \sum_{s, \sigma=1}^{S} \beta_s \beta_\sigma K_s K_\sigma w \right)
\]

\[
F_{s\sigma}(\beta) = w^T K_s^T K_\sigma w \quad F(w) = K_\beta^T K_\beta
\]

- Information Optimality Criteria: Maximize Information by minimizing variance of estimator.

- D-optimality:

\[
\beta \leftarrow \arg \min_{\beta} \left\{ -2 \log |K_\beta| \right\}
\]
• Maximum value of kernel combination parameter while varying alignment.

• Fisher MKL approach and an area of “learning”.
Conclusions

- FME and a lower bound for correlation of MKL responses - flat maxima.
- Ambiguity and BVC decompositions - accuracy & diversity.
- DoE approach for simple linear regression MKL case - Fisher information.

What’s next

- Fisher MKL for Classification.
- Analyzing the effect of sparsity [prior-regularization] and “localized” MKL?

Code

- [Variational Bayes & Sparse models]
Preliminary Analysis of MKL
Flat Maxima, Diversity and Fisher Information

Thank you for your attention

Theo Damoulas
Institute for Computational Sustainability
Computing and Information Science
Cornell University
Ithaca, NY, USA

Mark A. Girolami
Inference Research Group
Dept. of Computing Science & Dept. of Statistics
University of Glasgow
Scotland, UK

Simon Rogers
Inference Research Group
Department of Computing Science
University of Glasgow
Scotland, UK

Understanding Multiple Kernel Learning Methods
NIPS Workshop, Whistler, December 2009