Grace Wahba – Scholar, Mentor, Pioneer

Ming Yuan[1] Yoonkyung Lee[2]

[1]Department of Statistics
Columbia University

[2]Department of Statistics
The Ohio State University

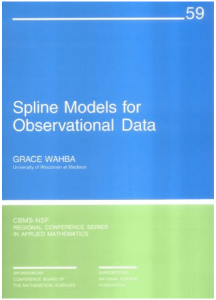
October 6, 2025
ISI World Statistics Congress, The Hague, Netherlands

Thursday Group

Key Contributions

Mother of Smoothing Splines

- Mathematical Foundation – RKHS
- Computation – Representer theorem
- Hyperparameter tuning – GCV
- Inferences – Gaussian process
- Multivariate – Smoothing Spline ANOVA



Reproducing Kernel Hilbert Spaces

Kimeldorf and Wahba (1971) connected RKHS to splines and developed the famous representer theorem.

- Provided an elegant and general mathematical treatment for fitting curves to noisy data
- The work also led to practical computational techniques
- Optimizing over an infinite dimensional RKHS can be reduced to optimizing over a finite dimensional linear subspace spanned by a certain fixed set of bounded linear functional

Generalized Cross Validation

Craven and Wahba (1978), Golub, Heath and Wahba (1979) developed the generalized cross-validation (GCV) technique to alleviate the computational cost of leave-one-out cross validation.

- Provided a deep analysis of GCV, proved its theoretical properties, and developed efficient code for its implementation
- Alleviated the computational burden while avoiding the estimation of other nuisance parameters such as the error variance
- GCV has become a standard method for tuning parameter selection, and it has had a huge impact on statistical practice and in machine learning algorithms

Smoothing Splines and Gaussian Process

Kimeldorf and Wahba (1970) established the equivalence between smoothing splines and nonparametric Bayesian inference based on a zero-mean Gaussian process as prior.

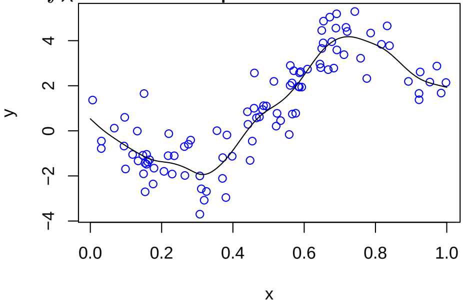
- Choice of polynomial splines led to improper priors
- Enabled Bayesian-style inference for smoothing splines
- Inspired extensive use of RKHS and Gaussian processes in Bayesian machine learning

Smoothing Splines

- $W_2[0,1] = \{f|\ f, f'$ absolutely continuous, and $\int_0^1 (f''(x))^2 dx < \infty\}$ (*Sobolev space*)
- Given data $\{(x_i, y_i)\}_{i=1}^n$, find $f \in W_2[0,1]$ minimizing

$$\sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda \int_0^1 (f''(x))^2 dx,$$

where $\lambda > 0$ is a smoothing parameter
- The solution $\hat{f}_\lambda$ is a cubic spline



Representer Theorem

Kimeldorf and Wahba (1971), *Some results on Tchebycheffian Spline Functions*

- View $W_2[0,1] = \text{span}\{1, x\} \oplus W_2^0[0,1] = \mathcal{H}_0 \oplus \mathcal{H}_1$ as a Hilbert space $\mathcal{H}$ equipped with an inner product such that $\int_0^1 (f''(x))^2 dx = \|P_1 f\|_{\mathcal{H}}^2$ and $f(x) = \langle f, K(x, \cdot)\rangle_{\mathcal{H}}$ with reproducing kernel K

- The minimizer $\hat{f}_\lambda \in \mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ of

$$\sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda \underbrace{\int_0^1 (f''(x))^2 dx}_{\|P_1 f\|_{\mathcal{H}}^2}$$

  is of the form

$$\hat{f}_\lambda(x) = (d_0 + d_1 x) + \sum_{i=1}^{n} c_i K_1(x_i, x).$$

- The solution belongs to a finite dimensional space!

Kernel (aka RKHS) Methods

- Model space: a reproducing kernel Hilbert space (RKHS) $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ on a domain $\mathcal{X}$ with kernel $K = K_0 + K_1$
- Given data, find a model $f \in \mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ minimizing

$$\sum_{i=1}^{n} L\left(y_i, f(x_i)\right) + \lambda \|P_1 f\|_{\mathcal{H}}^2$$

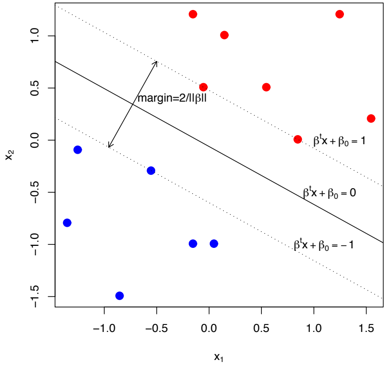
- When the null space $\mathcal{H}_0$ is spanned by $\{\phi_j\}_{j=1}^{m}$, the solution is of the form:

$$\hat{f}_\lambda(x) = \sum_{j=1}^{m} d_j \phi_j(x) + \sum_{i=1}^{n} c_i K_1(x_i, x)$$

- Need only to determine $d_1, \ldots, d_m$ and $c_1, \ldots, c_n$ in $\hat{f}_\lambda$

Support Vector Machines

- Vapnik and his collaborators invented support vector machines for binary classification in the '90s
- Maximum margin classifiers:
  - With $y_i \in \{\pm 1\}$, classifier: $\phi(x) = \text{sign}(f(x))$
  - Find $f(x) = \beta_0 + \beta^t x$ by maximizing $1/\|\beta\|$ subject to $y_i f(x_i) \geq 1$
  - Nonlinear generalization was done through the kernel trick



Support Vector Machines in RKHS

- SVMs find a "large-margin" discriminant $f(x) = \beta_0 + h(x)$ with $h \in \mathcal{H}$ by minimizing

$$\sum_{i=1}^{n} \underbrace{(1 - y_i f(x_i))_+}_{\text{Hinge loss}} + \lambda \|h\|_{\mathcal{H}}^2$$

- Margin maximization is the same as penalization with an RKHS norm!

- Hinge loss is not likelihood based, so how to interpret $f(x)$ statistically?

- Sensible extension to more than two classes?

Kernels in Machine Learning

- The success of SVMs generated a lot of interest in kernels
- Kernel methods of various forms proliferated
- Copious applications of the representer theorem followed
- The influence of the kernel based framework still continues!
  (e.g., arc-cosine kernel for infinitely wide NN, neural tangent kernel for deep learning)

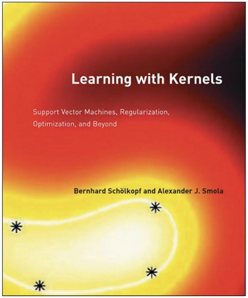


Figure: Book by Schölkopf and Smola

Conclusions

Her Advice to Young Statisticians

Wahba (2013), *Statistical Model Building, Machine Learning, and the Ah-Ha Moment*
Knutson (2017), *Breaking ground with Grace* (UW-News)

- "Keep your eyes open to synergies between apparently disparate fields."
- "Learn absolutely as much as you can about the subject matter of the data that you contemplate analyzing."
- "Find computer scientist friends."
- "If you want to do well, pick something you love and put your nose to the grindstone. It takes a lot of hard work, time and a certain amount of luck."

Thank you!

References I

Aronszajn, N. (1950).
Theory of reproducing kernel.
Transactions of the American Mathematical Society, 68:3337–404.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992).
A training algorithm for optimal margin classifiers.
In Haussler, D., editor, *Proceedings of the 5th Annual Workshop on Computational Learning Theory (COLT'92)*, pages 144–152. ACM Press.

Cho, Y. and Saul, L. (2009).
Kernel methods for deep learning.
NeurIPS, 22.

References II

Craven, P. and Wahba, G. (1978).
Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation.
Numerische mathematik, 31(4):377–403.

Golub, G. H., Heath, M., and Wahba, G. (1979).
Generalized cross-validation as a method for choosing a good ridge parameter.
Technometrics, 21(2):215–223.

Gu, C. (2002).
Smoothing Spline ANOVA Models.
Springer, New York.

References III

Jacot, A., Gabriel, F., and Hongler, C. (2018).
Neural tangent kernel: Convergence and generalization in neural networks.
NeurIPS, 31.

Kimeldorf, G. and Wahba, G. (1971).
Some results on Tchebycheffian spline functions.
Journal of Mathematical Analysis and Applications, 33(1):82–95.

Lee, Y., Lin, Y., and Wahba, G. (2004).
Multicategory support vector machines, theory, and application to the classification of microarray data and satellite radiance data.
Journal of the American Statistical Association, 99:67 – 81.

Wahba, G. (1990).
Spline Models for Observational Data.
Series in Applied Mathematics, Vol. 59, SIAM, Philadelphia.

References IV

Wahba, G. (1998).
Support vector machines, reproducing kernel Hilbert spaces, and randomized GACV.
In Schölkopf, B., Burges, C. J. C., and Smola, A. J., editors, *Advances in Kernel Methods: Support Vector Learning*, pages 69–87. MIT Press.

Wahba, G. (2002).
Soft and hard classification by reproducing kernel Hilbert space methods.
Proceedings of the National Academy of Sciences, 99:16524–16530.