Contents

Chapter 1

Interpolation Methods

20   Table - overview parameter: h: bandwidth, λ: penalty-factor, k: neighbors

21   categories

22   tuning parameter unbiased asympt. unbiased deg. of freedom continuous smooth (C^∞)

1.1   Setting

23

We are given data in the form of (x_i, Y_i) for $i = 1, \ldots, n)$. Assume that it can be represented by

$$Y_i = m(x_i) + \varepsilon_i,$$

where ε_i is some noise and $m : \mathbb{R} \to \mathbb{R}$ being some (non-parametric regression) function. If we assume that $\varepsilon_1, \ldots, \varepsilon_n$ i.i.d. with $\mathbb{E}[\varepsilon_i] = 0$ then

$$m(x) = \mathbb{E}[Y \mid x]$$

24   Different assumptions on m will lead to the following models:

1.2   Methods - Description

25

1.2.1   Kernel Regression

26

$$\mathbb{E}[Y \mid X = x] = \int_{\mathbb{R}} y f_{Y|X}(y \mid x) dy = \frac{\int_{\mathbb{R}} y f_{X,Y}(x, y) dy}{f_X(x)},$$

where $f_{Y|X}, f_{X,Y}, f_X$ denote the conditional, joint and marginal densities. We can estimate those with a kernel K by

$$\hat{f}_X(x) = \frac{\sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)}{nh}, \hat{f}_{X,Y}(x, y) = \frac{\sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) K\left(\frac{y - Y_i}{h}\right)}{nh^2}$$

Approximating the integral by a Riemann sum this yields the Nadaraya-Watson kernel estimator:

$$\hat{m}(x) = \frac{\sum_{i=1}^n K\left((x - x_i)/h\right) Y_i}{\sum_{i=1}^n K\left((x - x_i)/h\right)}$$

27   **Pros**: - can be assigned degrees of freedom (trace of the hat-matrix) - estimation of
28   the noise variance $\hat{\sigma}_\varepsilon^2$ (XXX c.f. CompStat 3.2.2)

29 **Cons**: - choice of kernel - if the $x \mapsto K(x)$ is not continuous, $\hat{m}$ isn't either - choice of
30 bandwidth, especially if x_i are not equidistant.

31 **Examples:** Normal, Box For local bandwidth selection see Brockmann et al. (1993)
32 XXX

33 1.2.2 Running Mean

34 1.2.3 Savitzky–Golay filter

For this section we refer to (Schafer, "What Is a Savitzky-Golay Filter?"). This technique
is used in signal processing and can be used to filter out high frequencies . But it can
also used for smoothing by filtering high frequency noise while keeping the low frequency
signal. First we choose a window size m. Then, for each point $j \in \{m, m+1, \ldots, n-m\}$
we fit a polynomial of degree k by:

$$\hat{y}_j = \min_{p \in P_k} \sum_{i=-m}^{m} (p(x_{j+i}) - y_{i+j})^2,$$

35 were P_k denotes the Polynomials of degree k over $\mathbb{R}$.

For equidistant points this can efficiently be calculated by

$$\hat{y}_j = \sum_{i=-m}^{m} c_i y_{j+i},$$

36 where the c_i are only dependent on the m and k and are tabulated in (XXXcitationXXX).

37 **Pros** - popular technique in signal processing - efficient calculation for equidistant
38 points

39 **Cons** - no natural way of how to estimate points which are not in the data.

40 1.2.4 Polynomial interpolation

41 1.2.5 Polynomial approximation

42 1.2.6 loess

43 1.2.7 Cubic Smoothing Splines

44 We interpolate with a function in C^2 (space of three time continuous differentiable func-
45 tions) which is defined piecewise by cubic polynomials. **Pros** Regression splines (B-
46 splines) XXXciteXXX Wood (2017) use a basis of the spline space (eg B-splines or j-th
47 cardinal basis) and fit the splines of degree k to approximate the data.

48 1.2.8 B-splines

from T. Lyche (T. Lyche)

$$S(x) = \sum_{j=0}^{n-1} c_j B_{j,k;t}(x)$$

$$B_{i,0}(x) = 1, \text{ if } t_i \leq x < t_{i+1}, \text{ otherwise } 0$$
$$B_{i,k}(x) = \frac{x-t_i}{t_{i+k}-t_i} B_{i,k-1}(x) + \frac{t_{i+k+1}-x}{t_{i+k+1}-t_{i+1}} B_{i+1,k-1}(x)$$

49 **Smoothing:** We can relax the constrain that we have to perfectly interpolate. Thus
50 we use the minimum number of knots[1] such that: $\sum_{i=1}^{n}(w(y_i - \hat{y}_i))^2 \leq s$

51 **Pros** - can be assigned degrees of freedom - extendable to "smooth" version - performs
52 also well if points are not equidistant

53 **Cons** - smoothing process does not translate well to a interpretation (unlike smoothing
54 splines) - choice of smoothing parameter s

55 ### 1.2.9   Natural Smoothing Splines

Let $\mathcal{F}$ be the Sobolev space (the space of functions of which the second derivative is
integrable). Then the unique[2] minimizer

$$\hat{m} := \arg\min_{f \in \mathcal{F}} \sum_{i=1}^{n} (Y_i - f(x_i))^2 + \lambda \int f''(x)^2 dx$$

56 is a natural[3] cubic spline.

57 **Pros:** - can be assigned degrees of freedom (trace of the hat-matrix) - efficient esti-
58 mation (closed form solution) - intuitive penalty (we don't want the function to be too
59 "wobbly" — change slopes) - performs also well if points are not equidistant - fixes the
60 Runge's phenomenon (fluctuation of high degree polynomial interpolation)

61 **Cons:** - choose λ

62 ### 1.2.10   Penalized Regression Splines

63 Intuition: similar as Natural Smoothing Splines, but we choose knots

64 ### 1.2.11   Kriging

65 — Diggle and Ribeiro (dig)

66 Idea / Justification / Kriging was developed in geostatistics to deal with autocorrelation of
67 the response variable at nearby points. By applying the notion that two spectral indices
68 which are (timewise) close should also take similar values we justify the application of
69 Kriging.

70 Definitions and Assumptions In the end we would like to fit a smooth Gaussian process to
71 the data. **Gaussian Process** $\{S(t) : t \in \mathbb{R}\}$ is a stochastic process if $(S(t_1), \ldots, S(t_k))$
72 has a multivariate Gaussian distribution for every collection of times $t_1, \ldots, t_k$. S can
73 be fully characterized by the mean $\mu(t) := E[S(t)]$ and its covariance function $\gamma(t, t') =$
74 $\mathrm{Cov}(S(t), S(t'))$

75 Assumption: We will assume the Gaussian process to be stationary. That is for $\mu(t)$ to be
76 constant in t and $\gamma(t, t')$ to depend only on $h = t - t'$. Thus, we will write in the following
77 only $\gamma(h)$.

78 Note that the process is also isotropic (i.e. $\gamma(h) = \gamma(\|h\|)$ since we are in a one-dimensional
79 setting and the covariance is symmetric.

---

[1]SciPy uses FITPACK and DFITPACK, the documentation suggests that smoothness is achieved by reducing the number knots used

[2]Strictly speaking it is only unique for $\lambda > 0$

[3]It is called natural since it is affine outside the data range ($\forall x \notin [x_1, x_n] : \hat{m}''(x) = 0$)

We also define the variogram of a Gaussian process as

$$V(h) := V(t, t+h) := \frac{1}{2} \text{Var}\left(S(t) - S(t+h)\right) = (\gamma(0))^2 (1 - \text{corr}(S(t), S(t+h)))$$

80   Ordinary Kriging Universal Kriging

Bibliography

Gaussian models for geostatistical data. In P. J. Diggle and P. J. Ribeiro (Eds.), *Model-Based Geostatistics*, pp. 46–78. Springer.

T. Lyche, K. M. Spline Methods.

Appendix A

Hi Mom