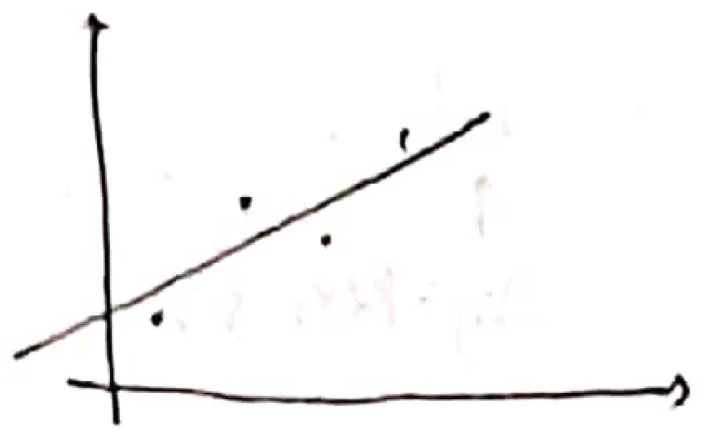
Recall: What is smoother? (or Regression).



$y = m(x) + e$

Find $m(x) = E(y \mid X = x)$　"smoother" $\hat{m}(X) = \begin{pmatrix} \hat{m}(x_1) \\ \vdots \\ \hat{m}(x_n) \end{pmatrix}$

Linear smoother $\hat{m}(X) = [*]\ y$.

$\hat{m} = S\ y$

e.g. kernel method, Loess

Lec 15 Splines: approximate $m$ is a simpler space $= \mathrm{Span}\{e_1, \dots, e_n\}$, where $\{e_i\}$ are basis (functions).

Def ① k阶样条：分段 k 次多项式，端点 (knots) 处 $0 \sim (k-1)$ 次导连续

它有基 truncated power basis: $1, x, \dots, x^k, (x - x_i)_+^k$, $i = 1, \dots, n$　共 $k+1+n$ 个. $= (k+1)(n+1) - kn$　→ natural splines 无

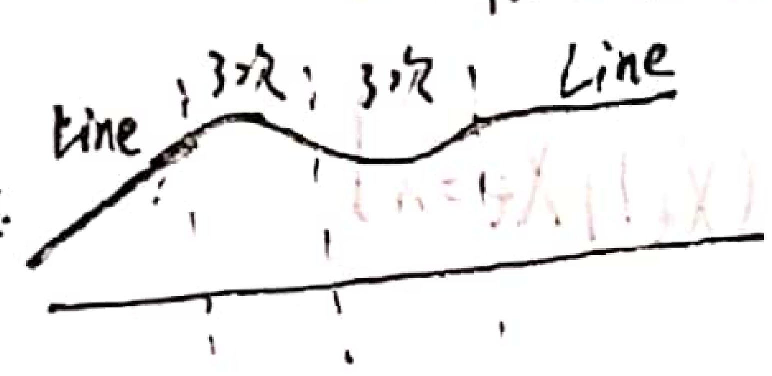
· 会算 df.（基的个数）.

② B-样条：一组 B样条基组成的空间，基的表达略.

③ k阶自然样条：内分段 k 次多项式，端点 $0 \sim (k-1)$ 次导连续，外两段 $\frac{k-1}{2}$ 次多项式

它有基 $N_1, \dots, N_n$　共 $n$ 个 $= (k+1)(n+1) - \frac{k+1}{2} \times 2 - kn = n$

e.g. 三阶自然样条 = 光滑样条




· Spline smoother: 设 $\hat{m}(x) = \sum_{j=1}^{n} \hat{\beta}_j e_j(x)$

(regression spline, 更广义地说)

$\hat{\beta} = \arg\min \|y - \hat{m}(X)\|^2 = \arg\min \|y - G\beta\|^2$, 其中 $G_{ij} = e_j(x_i)$. $\Rightarrow \hat{\beta} = (G^TG)^{-1}G^Ty$

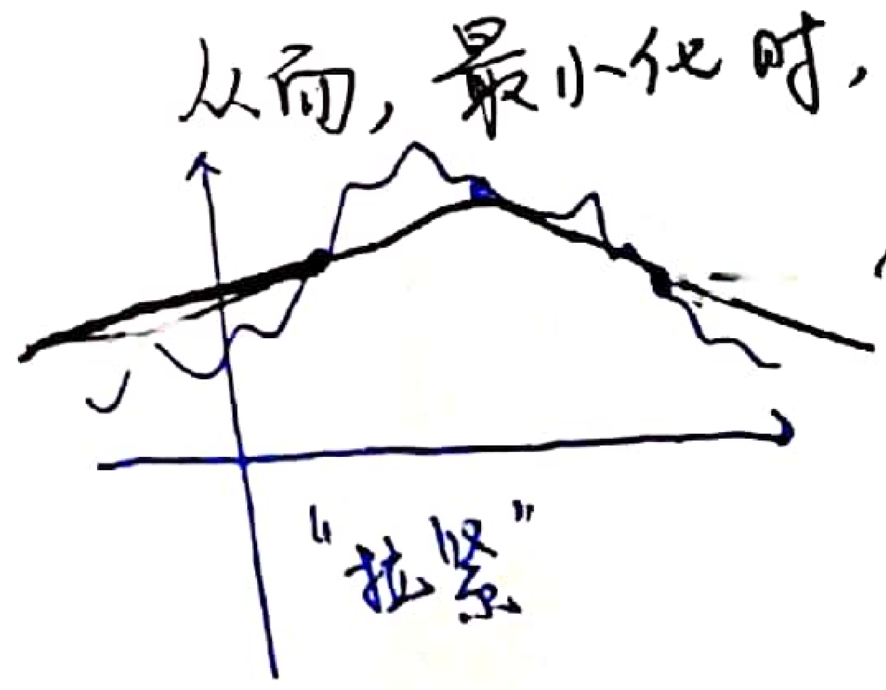
$\hat{m}(X) = G(G^TG)^{-1}G^Ty$.

· Smoothing splines: $\min\limits_{g\ 2次可导} \sum_{i=1}^{n}(y_i - g(x_i))^2 + \lambda \int_a^b (g^{(\frac{k+1}{2})}(t))^2 dt$　$a \le x_1 \le \cdots \le x_n \le b$.

则 minimizer is the k-th natural splines.

Thm (k=3 case) 固定插值，在自然样条时 $\int_a^b (g''(t))^2 dt$ 最小　☆ 会证明.

从而，最小化时，只须找 $\min\limits_{g\ 三次样条} \sum(y_i - g(x_i))^2 + \lambda \int (g''(t))^2 dt$，这时，$g = \sum_{j=1}^{n} \beta_j e_j$，则有




$\hat{\beta} = \arg\min\limits_{\beta} \|y - G\beta\|^2 + \lambda \beta^T \Omega \beta = (G^TG + \lambda\Omega)^{-1}G^Ty$.

类似 penalized LS.

其中 $G_{ij} = e_j(x_i)$, $\Omega_{ij} = \int_a^b e_j'' e_i'' dt$

$\hat{m}(X) = G(G^TG + \lambda\Omega)^{-1}G^Ty = (I + \lambda Q)^{-1}y$, $Q = (G^T)^{-1}\Omega G^{-1}$

若 $Q = \sum d_i u_i u_i^T$，则 $(I + \lambda Q)^{-1} = \sum \frac{1}{1 + \lambda d_i} u_i u_i^T$ $\Rightarrow \hat{m}(X) = \sum \left(\frac{u_i^T y}{1 + \lambda d_i}\right) \cdot u_i$

* RKHS 为 $g(x)=\sum_{i=1}^{n} c_i \phi_i(x)$（其中 kernel 为 $k=\sum \lambda_i \phi_i(x)\phi_i(z)$）

$$\hat{g} = \arg\min \sum (y_i - g(x_i))^2 + \lambda \|g\|_{\mathcal{H}}^2 = \arg\min \|y - KC\|^2 + \lambda C'KC = K(K+\lambda I)^{-1}y = (I+\lambda K^{-1})^{-1}y$$

↑ Hilbert space norm.

Lec 16 Splines: properties.

Def. $df(\hat{y}) = tr(S)$，其中 $\hat{y} = Sy$. 为 Smoother

$= \frac{1}{\sigma^2} tr(Cov(\hat{y}, y))$，其中 $e \sim N(0,\sigma^2)$

（或定义为 $2tr(S) - tr(SS')$，since $E(y-\hat{y})'(y-\hat{y}) = \sigma^2(n - 2tr(S) + tr(SS'))$.）

- 对于 smoothing splines，$df(\hat{g}) = \sum \frac{1}{1+\lambda d_j}$

- 进一步统计推断：CI，bias-correction，tests (Significant tests)

Lec 17 Applications:

- 变系数：$Y = a_0 + \sum_{j=1}^{p} a_j(U_j) X_j + \varepsilon_j$

方法：在 $u$ 处展开 $a_j(U_j)$，然后用 kernel method.（local linear estimator）

- 估计 Quantile（条件的）：$q_\tau(x) = \arg\min_{q} E[\rho_\tau(Y_t - q(X_t)) | X_t = x]$

方法 ①在 $x$ 处展开 $q_\tau(x)$，...

② 估出 CDF（条件的）$F_{Y|X}(y|x)$，用估计 CDF 的 quantile 作为 estimator

- Some tests

Lec 18. PLS

$Y_i = X_i^T \beta_0 + g(Z_i) + \varepsilon_i$，$i=1,\dots,n$

已知 + 未知；$1\times p$ 已知, linear；$q\times 1$ 未知.

i.i.d. $E(\varepsilon_i|X_i,Z_i)=0$，$E(\varepsilon_i^2|X_i,Z_i)=\sigma^2(X_i,Z_i)$

方法 1. 固定 $Z_i$，为线性模型 ⇒ 估 $\beta_0$（与 $Z$ 有关）

2. 估计 $Y_i|Z_i$ 的分布

3. Plug-in 2 的结果于 1.

另，要估 $\sigma^2(X_i, Z_i)$

进一步，统计推断：tests.

Lec 19 SIM

$E(y|x) = g(x^T\beta)$

未知（已知）

(SIR) $y_i = g(X_i^T\beta) + e_i$，$E(e_i|X_i)=0$.

未知

- Identification
- Ichimura's estimator:

类比 Leave-one-out CV 和 LS.

Lec 20 Additive Models