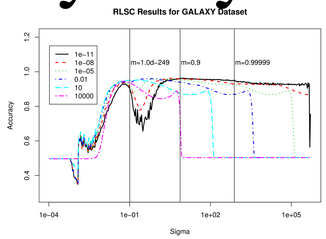
# Asymptotics of Gaussian Regularized Least Squares

R.A. Lippert and R.M. Rifkin

## A mystery ...




large $\sigma$ + small $\lambda$ = good classifier?!?!?

### Regularized Least Squares

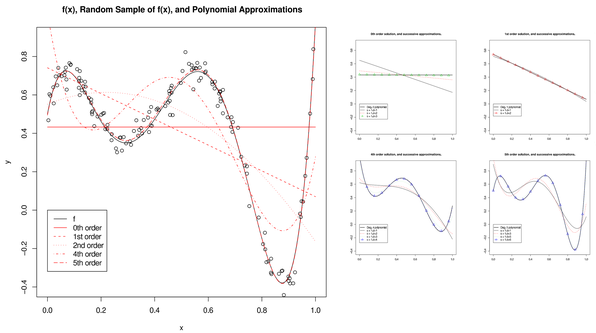
$$\begin{aligned} f(x_0) &= \sum_i c_i^{\sigma\lambda} K^\sigma(x_i, x_0) = K^\sigma[x, x_0] c^{\sigma\lambda} \\ c^{\sigma\lambda} &= (K^\sigma[x,x] + \lambda I)^{-1} y \\ f(x_0) &= y^t (K^\sigma[x,x] + \lambda I)^{-1} K^\sigma[x, x_0] \end{aligned}$$

*Key idea:* Cancellation of the $\sigma, \lambda$ terms in $f(x)$. (Derivation requires matrix tricks.)

## An answer ...

large $\sigma$ + small $\lambda$ = LS polynomial approximation

As $\sigma \to \infty$ and $\lambda \propto \sigma^{-(2p+1)}$, we see degree $p$ approximation.




## More generally ...

### Asymptotic polynomial approximation for convex loss

$$f(x) = \sum_{i=1}^n c_i K(x_i, x) \Rightarrow f(x) = \sum_I q_I x^I \quad \text{as } \sigma, \lambda \to \infty, 0$$

### Tikhonov regularization

$$f = \text{argmin}_{f \in \mathcal{H}} \left\{ \lambda ||f||_K^2 + \sum_i L(f(x_i), y_i) \right\}$$

### Regularization term turns into a constraint

$$(C\sigma^{-2p})||f||_{K^\sigma}^2 \Rightarrow \begin{cases} 0 & f \text{ is a polynomial of degree p} \\ \infty & \text{else} \end{cases} = C||f||_{\mathcal{P}_p}^2$$

(Actually, under certain conditions, it turns into a constraint plus a a residual regularization. We're ignoring that here.)

### And thus...

$$\text{argmin}_{f \in \mathcal{H}} \left\{ C\sigma^{-2p}||f||_K^2 + \sum_i L(f(x_i), y_i) \right\} \to \text{argmin}_f \left\{ C||f||_{\mathcal{P}_p}^2 + \sum_i L(f(x_i), y_i) \right\}$$

a uniqueness qualification

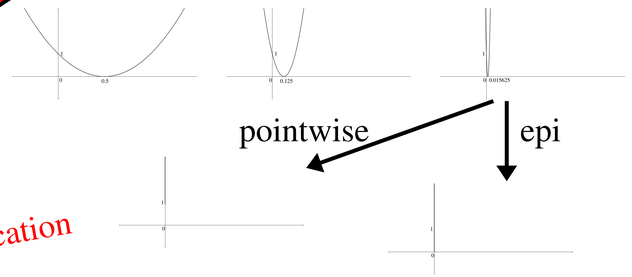
what kinds of kernels?

a tricky limit to take

Works on a general class of positive definite kernels:
$K(x, x') = \sum_{IJ} M_{IJ} x^I x'^J$,
where $M_{IJ}$ is spd on all finite diagonal blocks.
e.g. $K(x, x') = \sum_i (x \cdot x')^i \phi_i(x) \phi_i(x')$ with $\phi_i(0) \neq 0$.

### limits of minimizers versus minimizers of limits

$f \Rightarrow g$ means *epigraphical* not *pointwise* convergence




helpful resource: Variational Analysis, Rockafellar and Wets

**Theorem:**
If $G_\sigma \Rightarrow G$ and argmin$G_\sigma$, $G$ are unique, then argmin$G_\sigma \to$ argmin$G$.

### Proof sketch in 1D:

The real proof needs to

- handle the infinite matrices from the expansions rigorously
- use more complicated indexing for $n$-dimensions

**Lemma:** $Z_1(s), Z_2(s) \to 0$ as $s \to 0$ and $\begin{pmatrix} A & B^t \\ B & C \end{pmatrix}$ is spd, let

$$M_s = \begin{pmatrix} Z_1(s)^{-1} & 0 \\ 0 & Z_2(s) \end{pmatrix} \begin{pmatrix} A & B^t \\ B & C \end{pmatrix} \begin{pmatrix} Z_1(s)^{-1} & 0 \\ 0 & Z_2(s) \end{pmatrix}.$$

If $Q_s(y) = y^t M_s^{-1} y$ then $Q_s \Rightarrow Q$,

$$Q(y) = \begin{cases} 0 & y = \begin{pmatrix} * \\ 0 \end{pmatrix} \\ \infty & y = \begin{pmatrix} * \\ \neq 0 \end{pmatrix} \end{cases}.$$

**Theorem:** Let $K^\sigma(x, x') = \sum_{ij} M_{ij} \frac{x^i}{\sigma^i} \frac{x'^j}{\sigma^j}$, where $M$ is spd.
For $p > 0$ (non-integer), as $\sigma \to \infty$, $\sigma^{-2p}||\cdot||_{K^\sigma}^2 \Rightarrow Q(\cdot)$,

$$Q(f) = \begin{cases} 0 & f(x) = \sum_{i<p} \beta_i x^i \\ \infty & \text{else} \end{cases}.$$

**Proof:** If $f = \sum_{i=0}^\infty \beta_i x^i$, then

$$\begin{aligned} \sigma^{-2p}||f||_{K^\sigma}^2 &= \beta^t \left( \begin{pmatrix} \sigma^p & 0 & \cdots \\ 0 & \sigma^{p-1} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix} M \begin{pmatrix} \sigma^p & 0 & \cdots \\ 0 & \sigma^{p-1} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix} \right)^{-1} \beta \\ &= \beta^t \left( \begin{pmatrix} Z_1(\sigma^{-1})^{-1} & 0 \\ 0 & Z_2(\sigma^{-1}) \end{pmatrix} M \begin{pmatrix} Z_1(\sigma^{-1})^{-1} & 0 \\ 0 & Z_2(\sigma^{-1}) \end{pmatrix} \right)^{-1} \beta \end{aligned}$$

and **lemma** applies.

### A similar phenomenon first seen with SVMs

Compare this to Keerthi and Lin's 2003 result on SVMs: found linear SVMs when $\lambda \propto \sigma^{-2}$.
Something more general is happening ...