SVR and GMR

Comparison

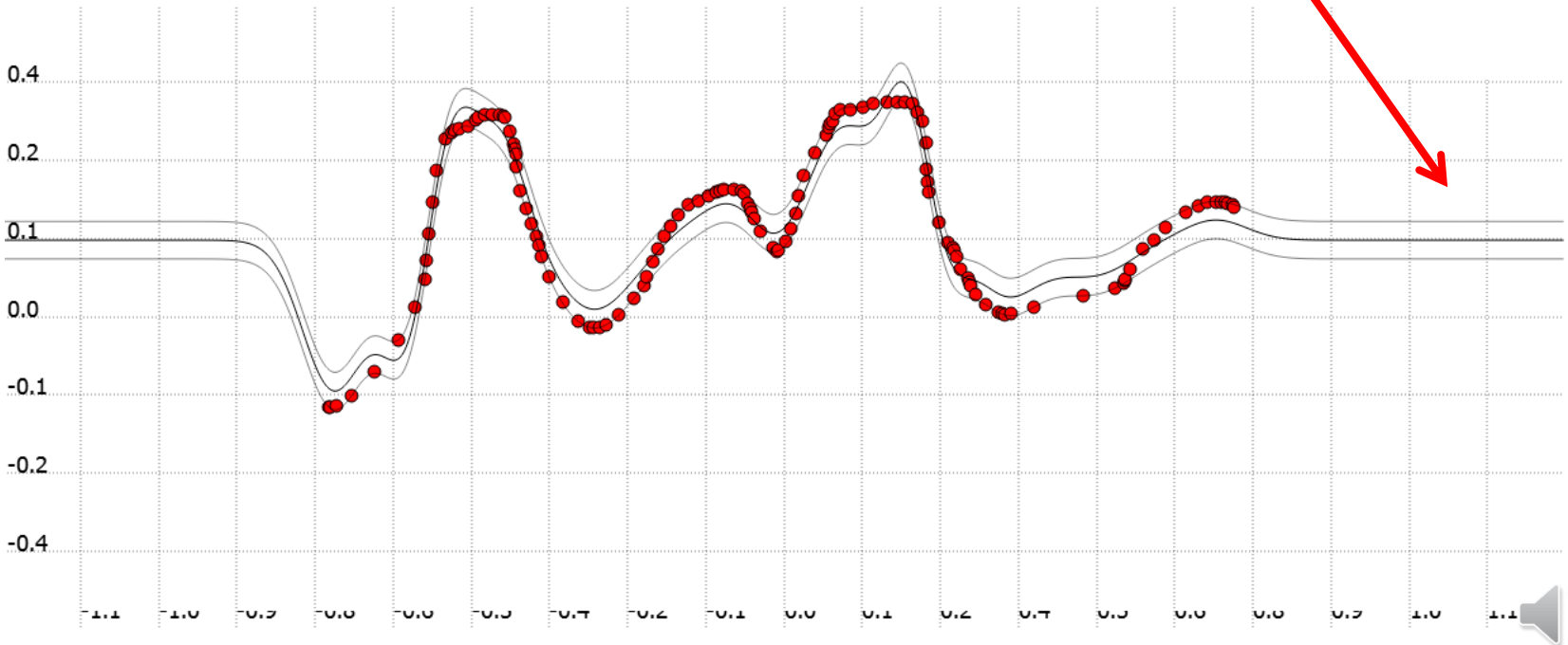
Comparison Across Methods

SVR Generalization – prediction away from datapoints

$$y = \sum_{i=1}^{M} \left(\alpha_i^* - \alpha_i \right) k\left(x^i, x \right) + b$$

$$\simeq 0$$

SVR predicts y=b away from datapoints

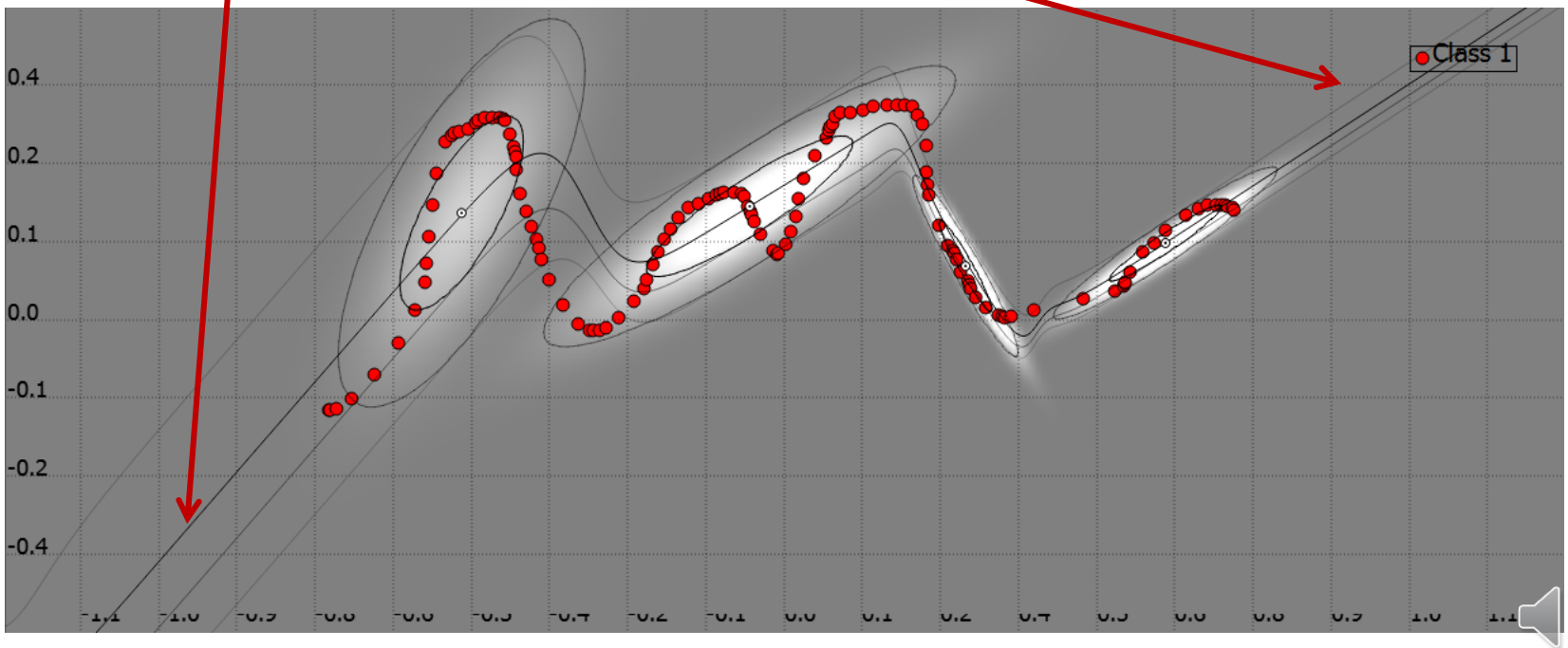


Comparison Across Methods

GMR Generalization – prediction away from datapoints

$$y = \sum_{k=1}^{K} \beta_k(x) \cdot \tilde{\mu}^k(x) \qquad \text{with} \quad \beta_k(x) = \frac{\alpha_k p\left(x; \mu^k, \Sigma^k\right)}{\sum_{j=1}^{K} \alpha_j p\left(x; \mu^j, \Sigma^j\right)}$$

GMR predicts the trend away from data

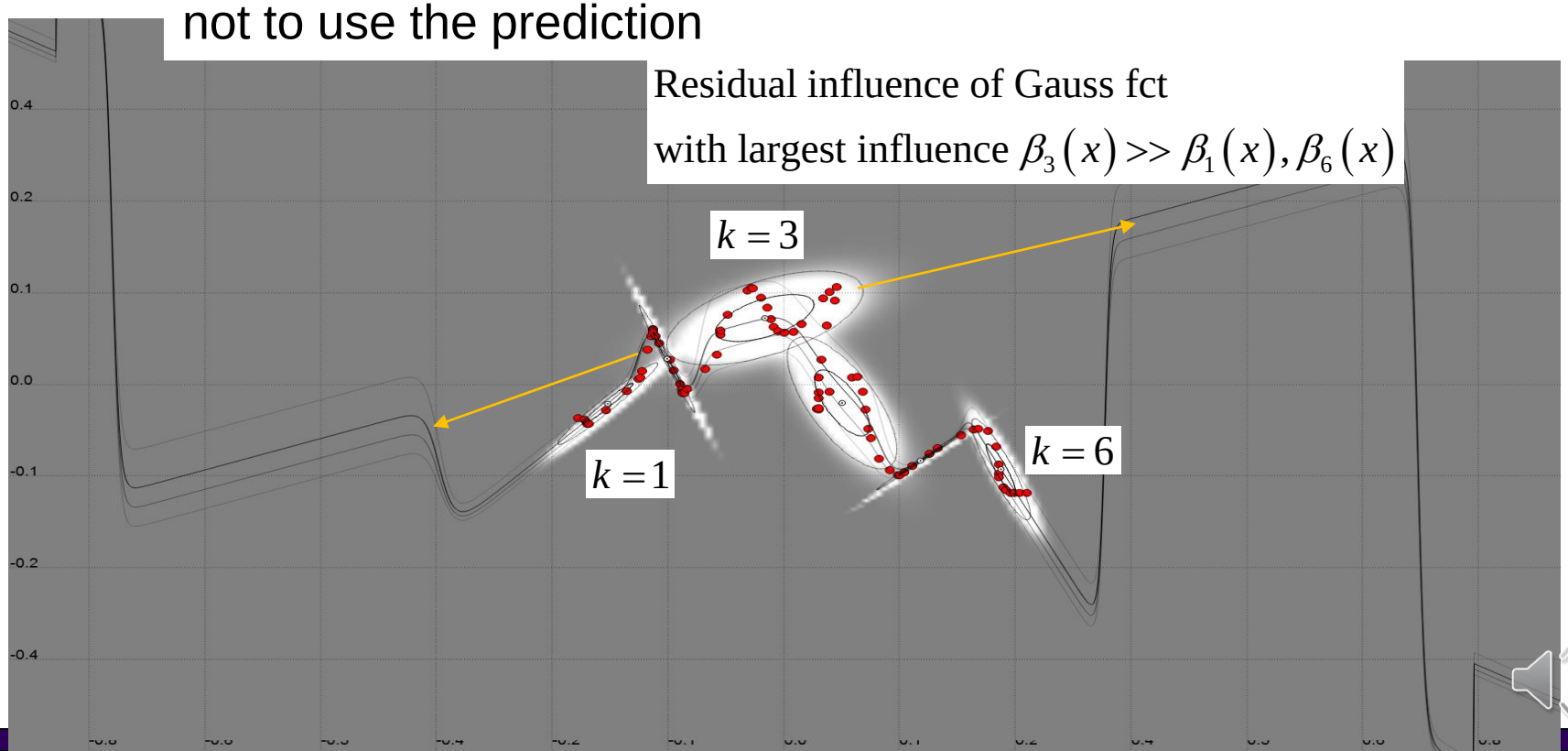


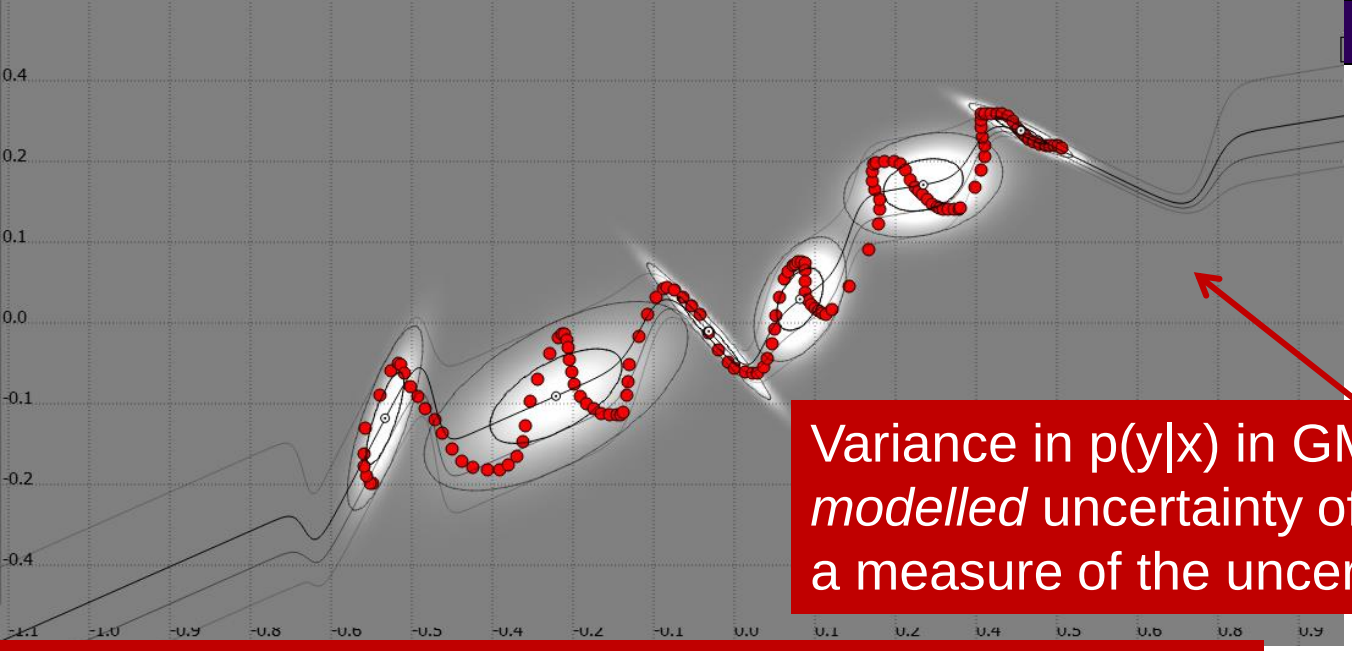
Comparison Across Methods

Generalization – prediction away from datapoints

The prediction away from the datapoints is affected by all regressive models. It may become meaningless!

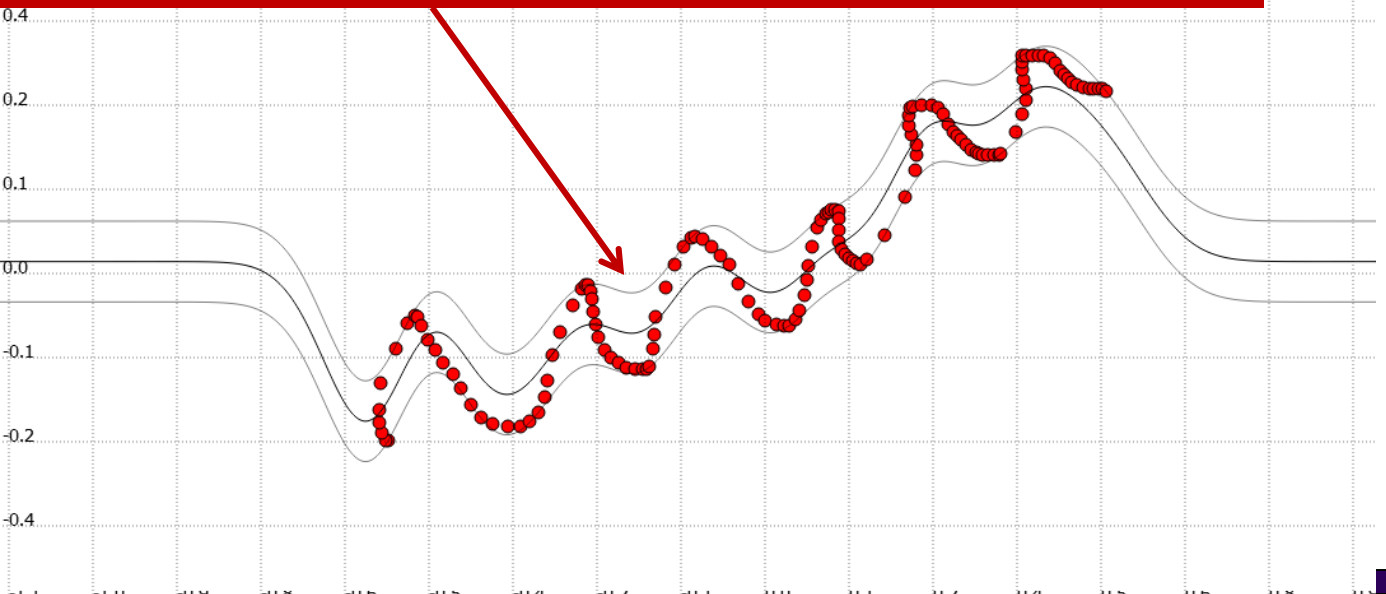
Use the likelihood of the model to determine whether it is safe or not to use the prediction





Variance in p(y|x) in GMR represents the *modelled* uncertainty of the value of y. It is not a measure of the uncertainty of the model.

Variance in SVR represents the epsilon-tube, the uncertainty around the predicted value of y. It does not represent uncertainty of the model either!



SVR, GMR: Similarities

SVR and GMR end up with a similar regressive model

SVR Solution

$$y = f(x) \pm \varepsilon$$

GMR Solution

$$\hat{y} = E\left\{p\left(y \mid x\right)\right\}$$

$$\Rightarrow y = f\left(x\right) \pm \varepsilon$$

$$\varepsilon \sim N\left(0, \mathrm{var}(p(y \mid x))\right)$$

SVR, GMR: Similarities

SVR and GMR compute a weighted combination of local predictors

SVR Solution

$$y = \sum_{i=1}^{M} \left(\alpha_i^* - \alpha_i \right) k\left(x^i, x \right) + b$$

GMR Solution

$$y = \sum_{i=1}^{K} \beta_i(x) \cdot \tilde{\mu}^i(x)$$

Both separate input space into regions modeled by Gaussian distributions (true only when using Gaussian/RBF kernels for SVR).

Model computed locally (locally weighted regression)!

SVR, GMR: Differences

GMR allows to predict multi-dimensional outputs, while SVR can predict only a uni-dimensional output y.

GMR starts by computing $p(y, x)$

$\Rightarrow$ can compute $p(x \mid y)$

x, y can have arbitrary dimensions

$\Rightarrow$ SVR cannot predict x from y.

SVR, GMR: Differences

SVR and GMR do not optimize the same objective function
→ They often find different solutions.

- SVR:
  - minimizes the error through convex optimization
  → ensured to find the *optimal* estimate; but not unique solution
  - usually finds a nm of models <= nm of datapoints (support vectors)

- GMR:
  - learns p(x,y) through maximum likelihood
  → finds *local* optimum
  - starts with a low nm of models << nm of datapoints

Hyperparameters of SVR, GMR

SVR and GMR depend on hyperparameters that need to be determined beforehand. These are:

- SVR
  - choice of error margin ε and penalty factor *C.*
  - choice of kernel and associated kernel parameters
- GMR:
  - choice of the number of Gaussians
  - choice of initialization (affects convergence to local optimum)

The hyperparamaters can be optimized separately; e.g. the nm of Gaussians in GMR can be estimated using BIC; the kernel parameters of SVR can be optimized through grid search.

Conclusion

No easy way to determine which regression technique fits best your problem

Training

SVR

Convex optimization
(SMO solver)
Parameters grow O(M*N)

GMR

EM, iterative technique,
needs several runs
Parameters grow $O(K*N^2)$

Testing

GMR

Grows *O(K)*

SVR

Grows
O(number of SV)
Few SV - Small fraction
of original data

M: number of datapoints; N: Dimension of data; K: Number of Gauss Functions in GMM model