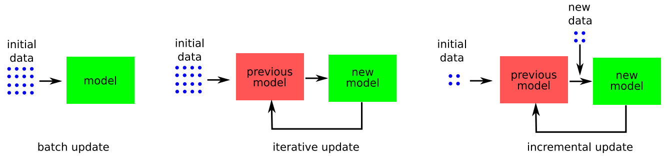
Regression

5. Incremental and iterative methods

Olivier Sigaud

Sorbonne Université
http://people.isir.upmc.fr/sigaud

Batch, Iterative, Incremental

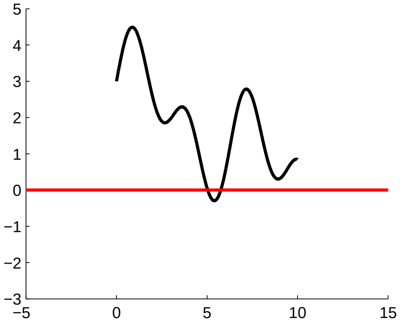


- So far, we have focused on batch regression
- Iterative methods: you improve a model through steps, using the same data at all steps
- Incremental methods: you improve a model through steps, with additional data at each step
- Incremental implies more than iterative
- Incremental approach to regression: reveal the batch data through steps

Limiting regression cost

- Solving $\boldsymbol{\theta}^* = (\mathbf{G}^\intercal\mathbf{G})^{-1}\mathbf{G}^\intercal\mathbf{y}$ (or any variant) requires inverting $(\mathbf{G}^\intercal\mathbf{G})$
- That is in $O(N^3)$ in the number of dimensions
- For large linear architectures or deep neural networks, this is too expensive
- Incremental solutions:
  - Complexity can be reduced to $O(N^2)$ by using the Sherman-Morrisson formula $\rightarrow$ incremental update of the inverse,
  - But sensitive to rounding errors.
  - Numerically more stable option: updating the Cholesky factor of the matrix using the QR algorithm.
  - In the linear case: Recursive Least Squares

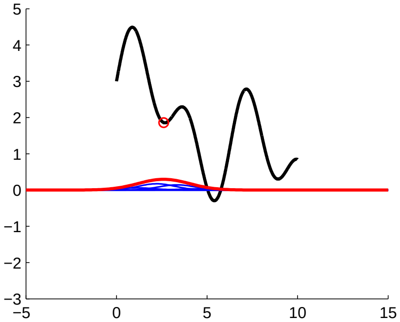
Radial Basis Function Networks (Illustration)



- Obtained using some incremental approach

Cybenko, G. (1989) Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)*, 2(4):303–314

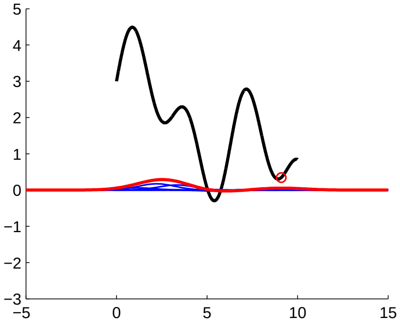
Radial Basis Function Networks (Illustration)



- Obtained using some incremental approach

Cybenko, G. (1989) Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)*, 2(4):303–314

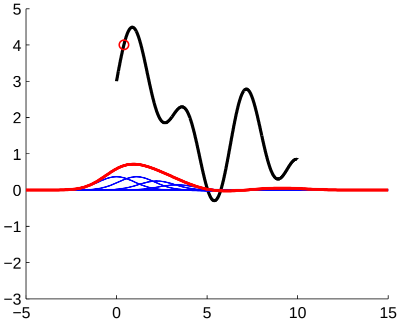
Radial Basis Function Networks (Illustration)



- Obtained using some incremental approach

Cybenko, G. (1989) Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)*, 2(4):303–314

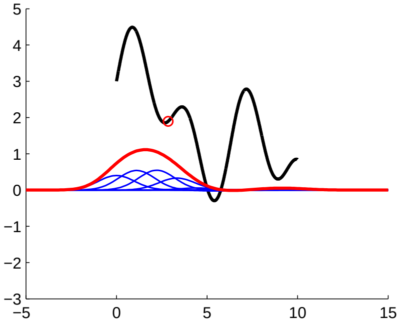
Radial Basis Function Networks (Illustration)



- Obtained using some incremental approach

Cybenko, G. (1989) Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)*, 2(4):303–314

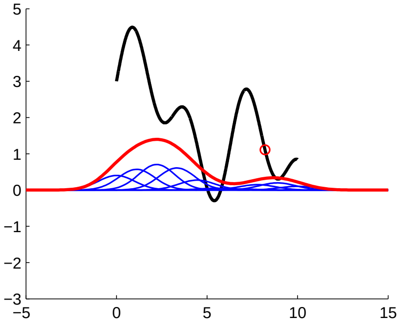
Radial Basis Function Networks (Illustration)



- Obtained using some incremental approach

Cybenko, G. (1989) Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)*, 2(4):303–314

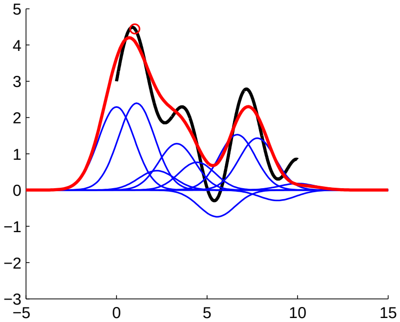
Radial Basis Function Networks (Illustration)



- Obtained using some incremental approach

Cybenko, G. (1989) Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)*, 2(4):303–314

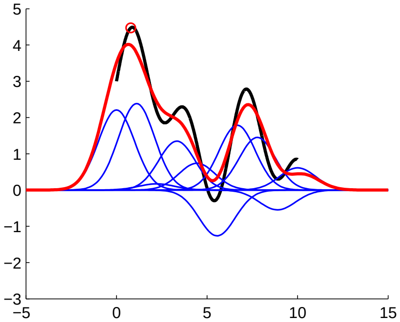
Radial Basis Function Networks (Illustration)



- Obtained using some incremental approach

Cybenko, G. (1989) Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)*, 2(4):303–314

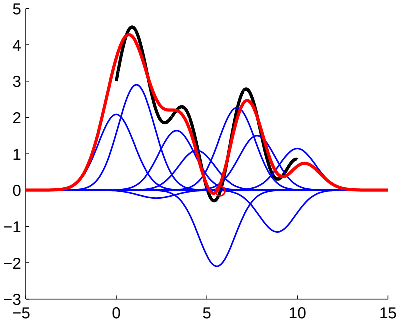
Radial Basis Function Networks (Illustration)



- Obtained using some incremental approach

Cybenko, G. (1989) Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)*, 2(4):303–314

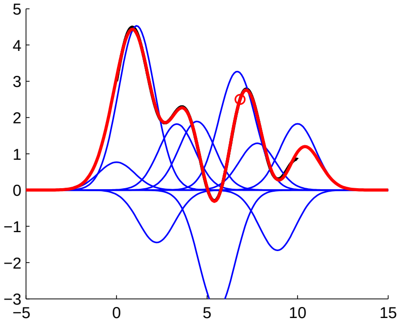
Radial Basis Function Networks (Illustration)



- Obtained using some incremental approach

Cybenko, G. (1989) Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)*, 2(4):303–314

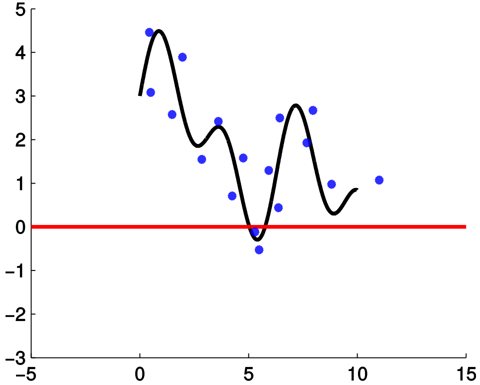
Radial Basis Function Networks (Illustration)



- Obtained using some incremental approach

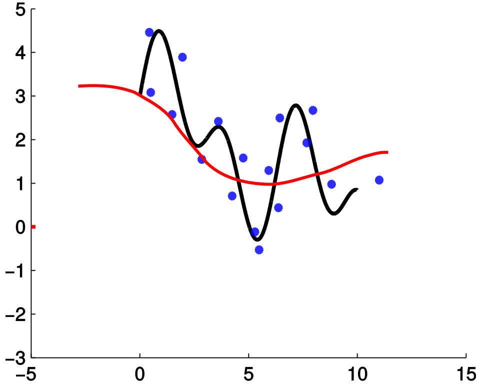
Cybenko, G. (1989) Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)*, 2(4):303–314

Risk of overfitting



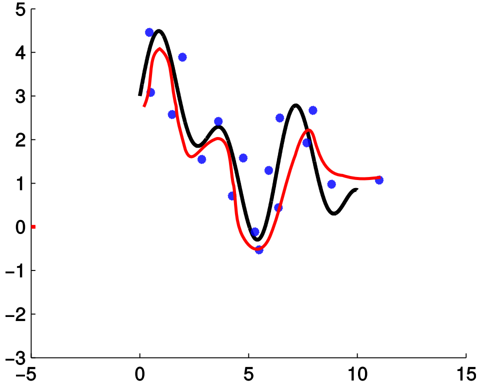
- In the previous simulation, measurements were noiseless
- What if we get noisy measurements?
- The risk is overfitting to data
- The question is when to stop

Risk of overfitting



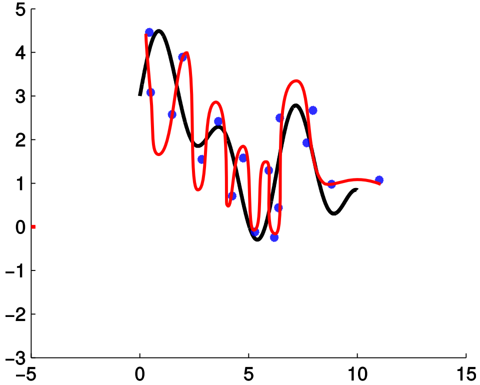
- In the previous simulation, measurements were noiseless
- What if we get noisy measurements?
- The risk is overfitting to data
- The question is when to stop

Risk of overfitting



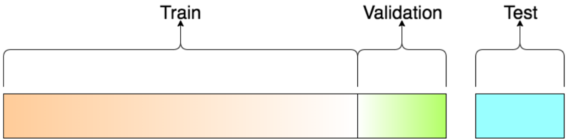
- In the previous simulation, measurements were noiseless
- What if we get noisy measurements?
- The risk is overfitting to data
- The question is when to stop

Risk of overfitting



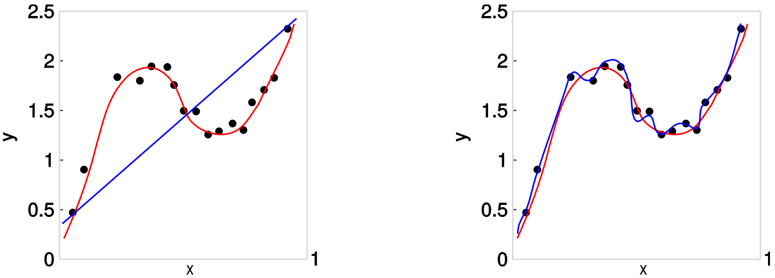
- In the previous simulation, measurements were noiseless
- What if we get noisy measurements?
- The risk is overfitting to data
- The question is when to stop

Fighting overfitting: Standard methodology



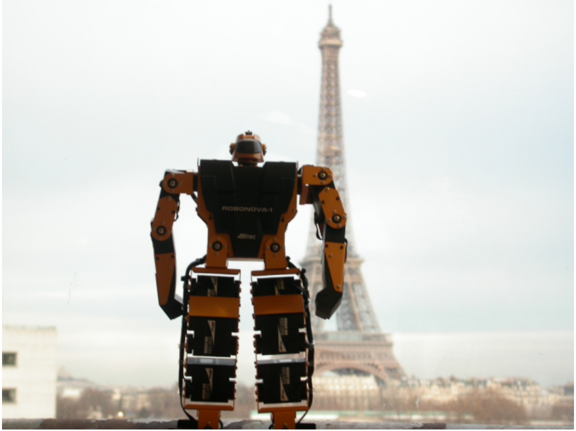
- Split the training set into training set and validation set
- Train on the training set
- Evaluate on the validation set
- When performance stops improving, stop iterating
- Should generalize well to the test set

Overfitting: influence of the model



- A model with more parameters is more prone to overfitting
- If the model does not have enough free parameters:
  - The training error may not go down towards 0
  - The validation performance may stop improving
- → Select a rich enough model...
- ...and use the validation trick to decide when to stop
- (Deep) neural networks are particularly rich models
- Training them with gradient descent in the next class

Any question?



Send mail to: `Olivier.Sigaud@upmc.fr`

Cybenko, G. (1989).
Approximation by superpositions of a sigmoidal function.
Mathematics of Control, Signals, and Systems (MCSS), 2(4):303–314.