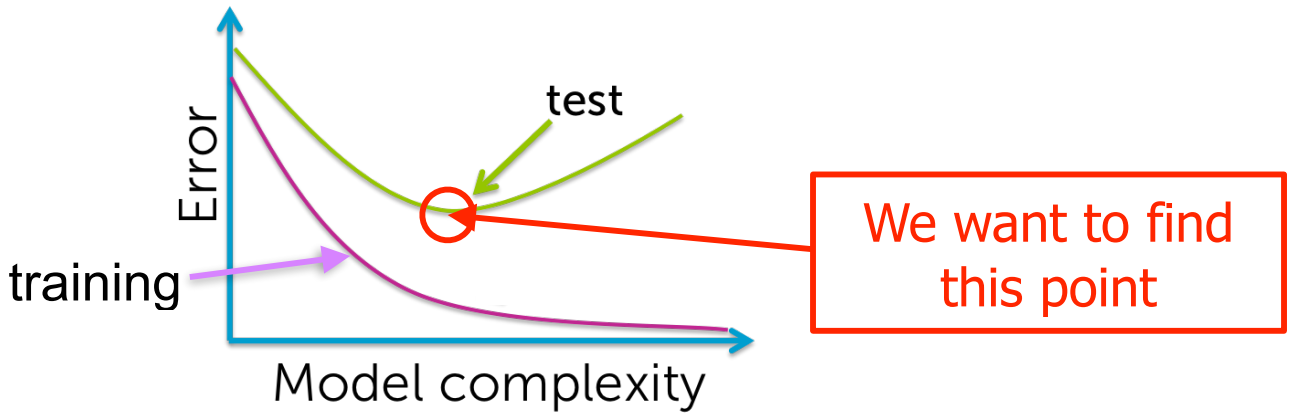
Validation

Pascal Fua
IC-CVLab

Model and Parameter Selection

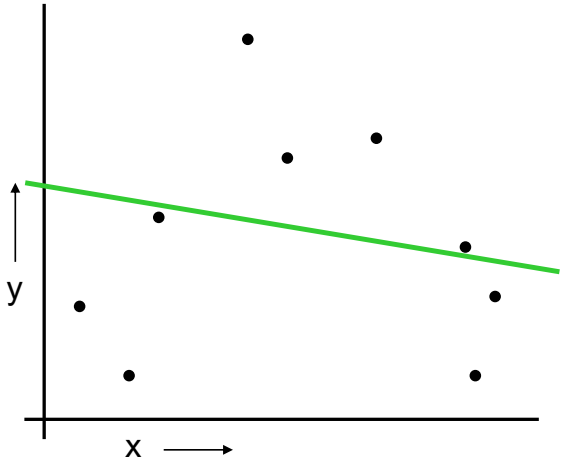


- Without access to the test data, how do we choose the right complexity and parameters for the model?

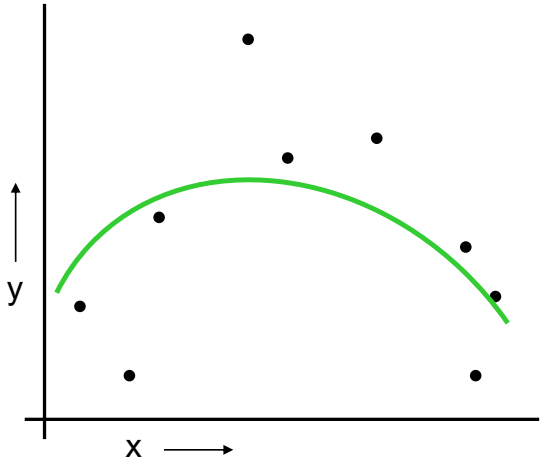
➡ Use a cross-validation technique.

Cross-validation: Toy example

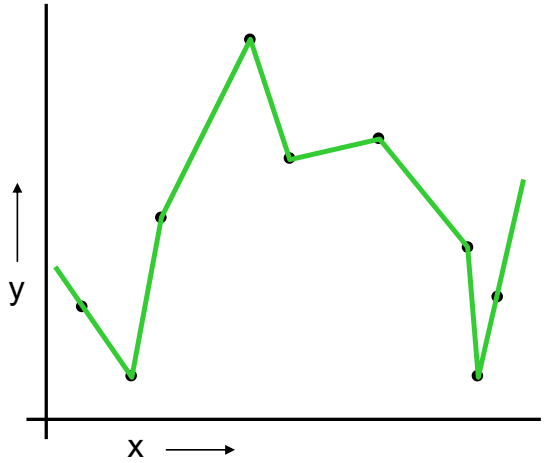
Linear function



Quadratic function

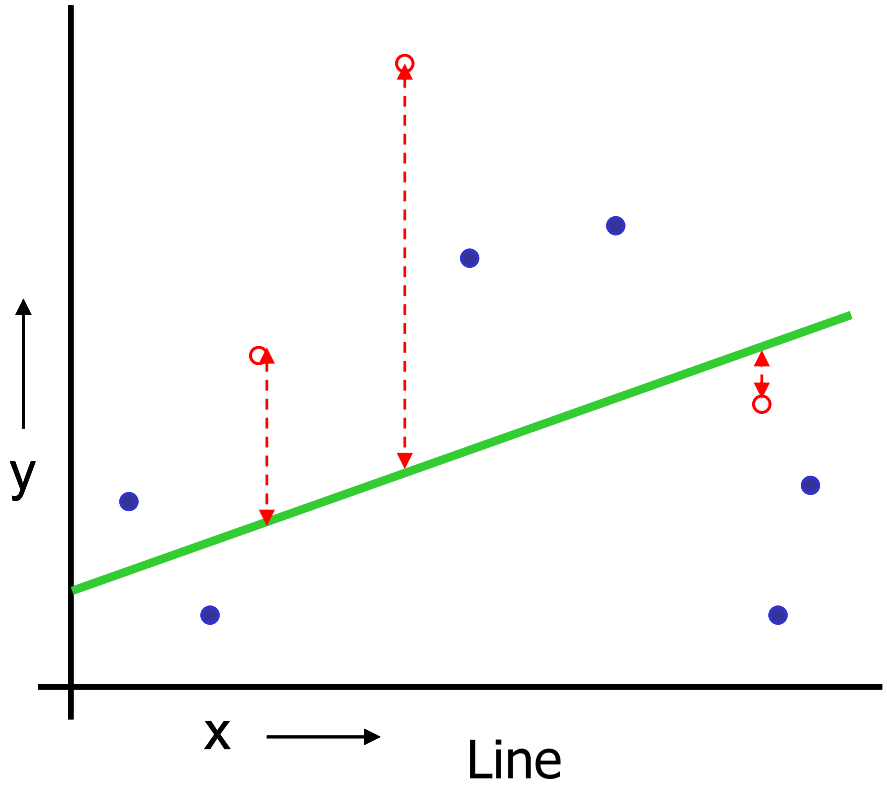


Piecewise linear function



Which is best?

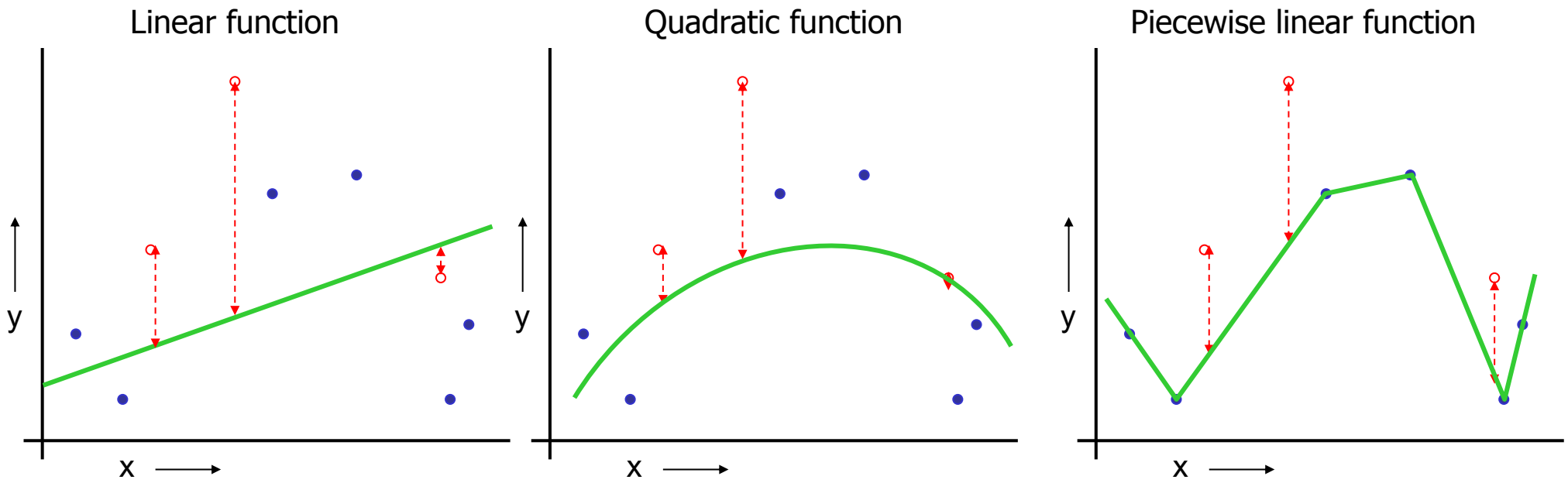
Validation Set Method



Mean squared error: 2.4

1. Randomly choose, e.g., 30% of the data to form a validation set
2. The remainder then acts as your new training set
3. Perform regression on this new training set only
4. Estimate the performance on the test data using the validation set

Validation Set Method



Mean squared error: 2.4

Mean squared error: 0.9

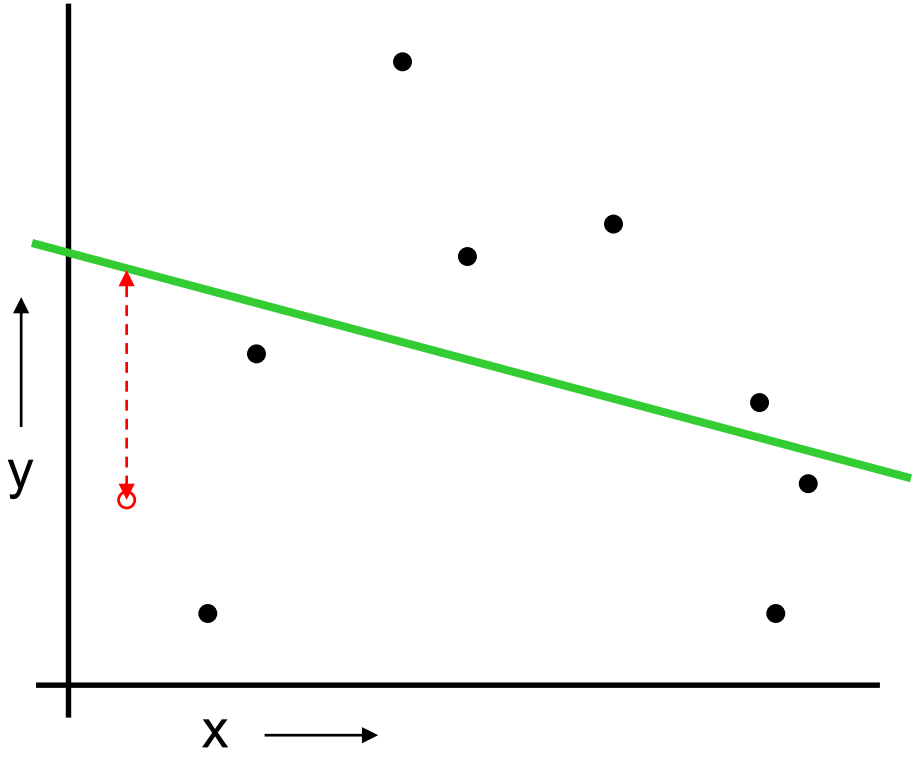
Mean squared error: 2.2

Validation Set Method

- **Strengths:**
  - Very simple
  - Reasonably effective

- **Drawbacks:**
  - Wastes data, the best model estimate was obtained using only a fraction of the data
  - If we don't have much data, the validation set might be lucky or unlucky

Leave-One-Out Cross Validation

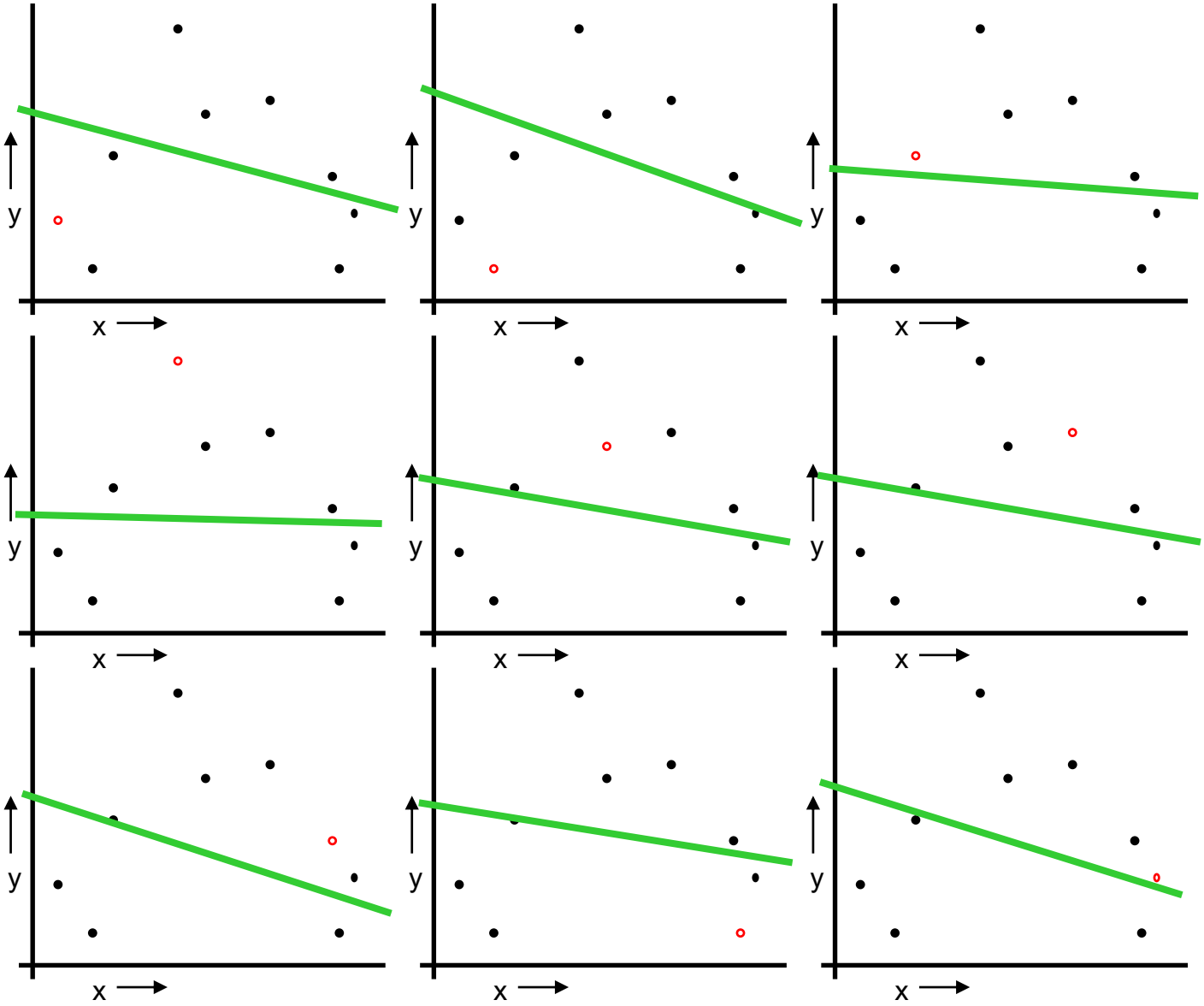


For $i = 1, \ldots, N$

1. Let (x_i, y_i) be the i^{th} sample
2. Temporarily remove (x_i, y_i) from the training set
3. Perform regression on the remaining $N - 1$ samples
4. Compute the error on (x_i, y_i)

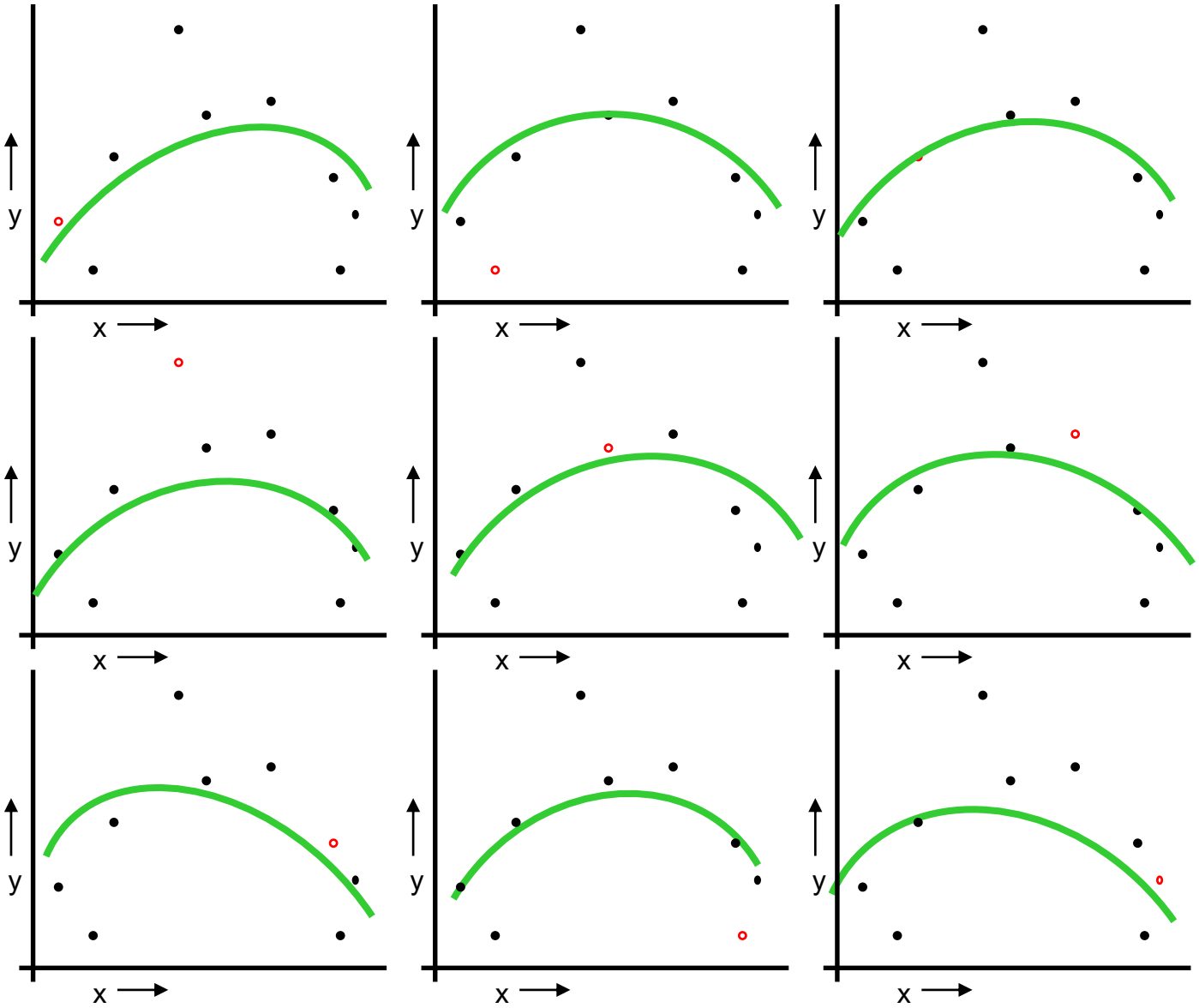
When done for all points, report the average error.

Leave-One-Out Cross Validation



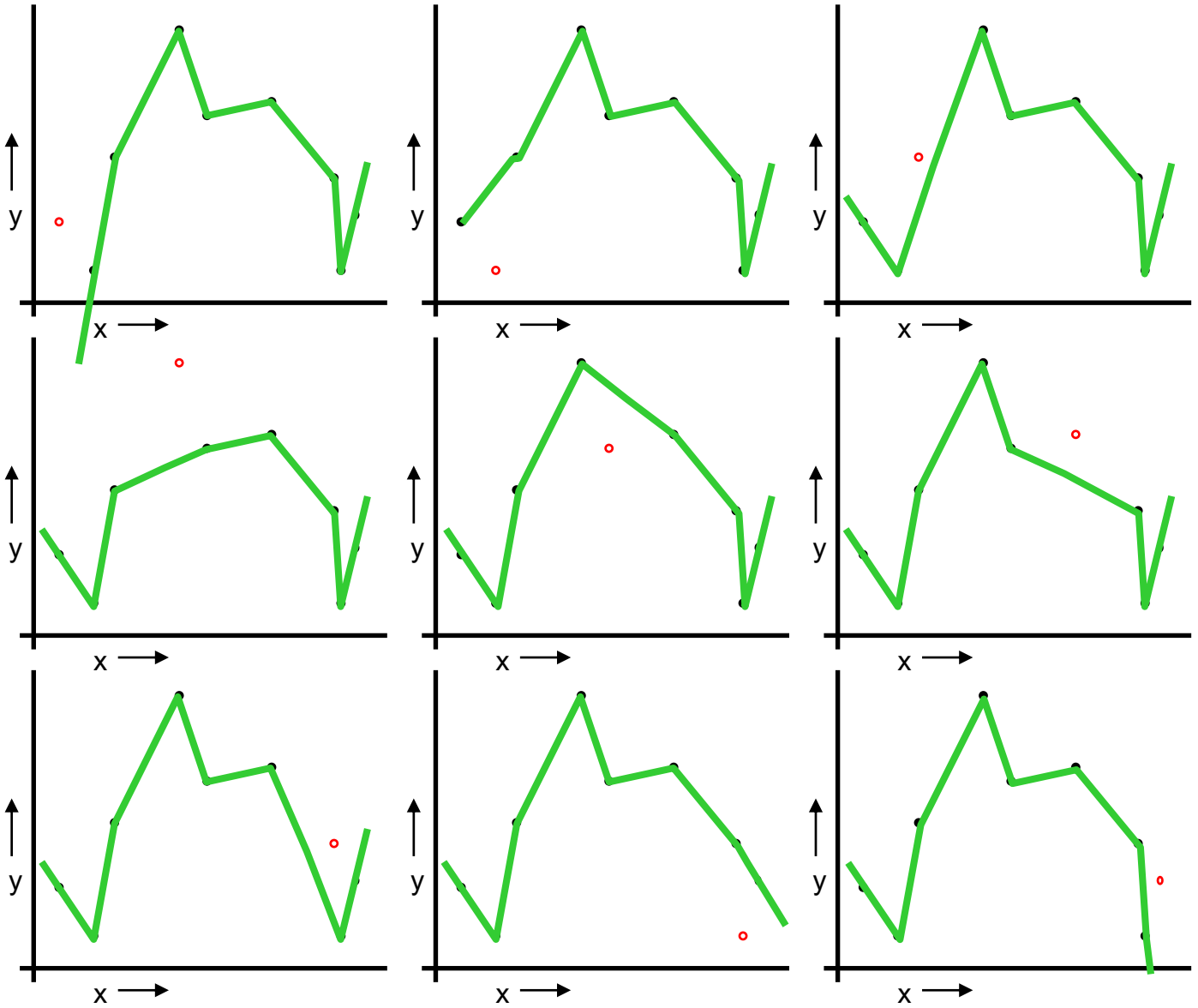
Mean squared error: 2.12

Leave-One-Out Cross Validation



Mean squared error: 0.962

Leave-One-Out Cross Validation



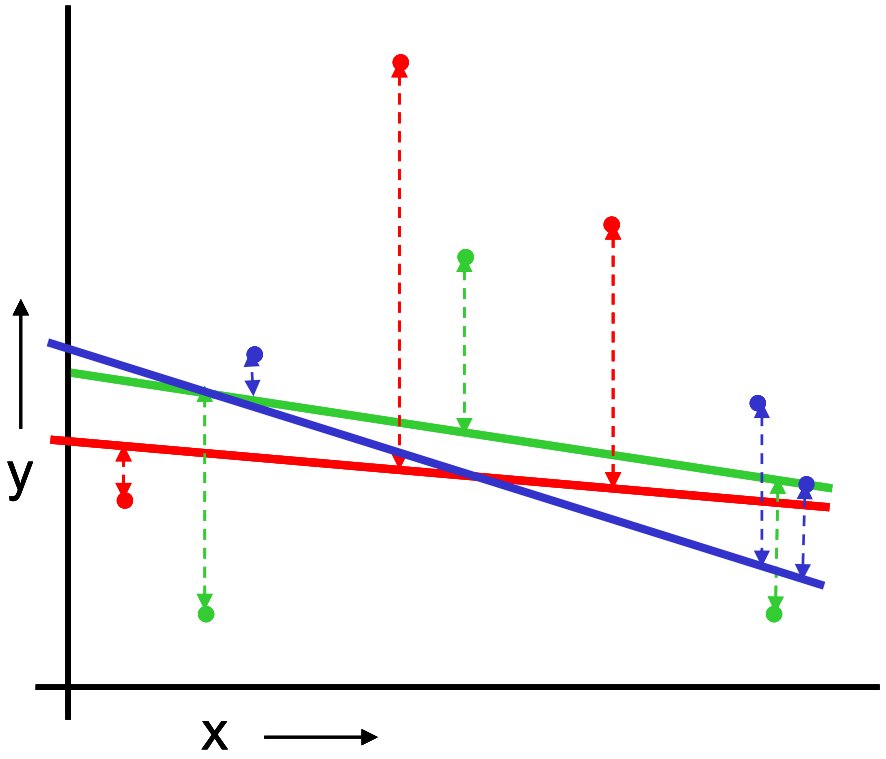
Mean squared error: 3.33

What Kind of Cross Validation?

	Downside	Upside
Validation-set	Potentially unreliable	Cheap
Leave-one-out	Expensive	Doesn't waste data

Can we get a god compromise?

k-Fold Cross Validation



Mean squared error: 2.05

Randomly split the dataset into k partitions (here $k = 3$, shown with different colors)

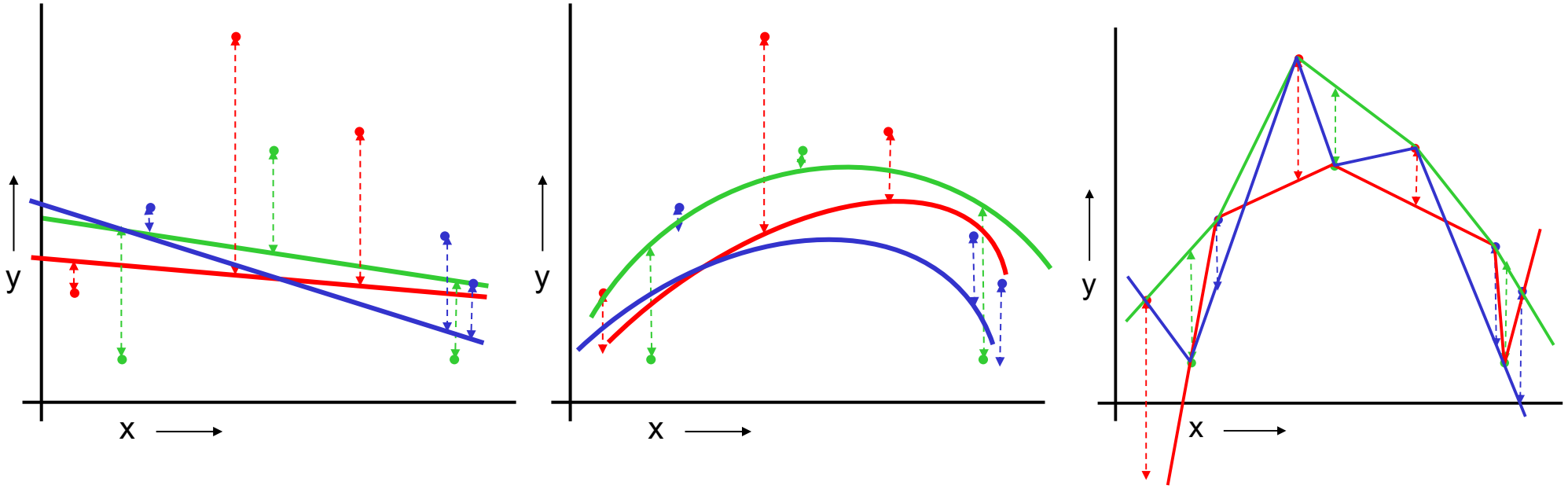
For the red partition, train on all the points not in the red partition. Compute the errors on the red points.

For the green partition, train on all the points not in the green partition. Compute the errors on the green points.

For the blue partition, train on all the points not in the blue partition. Compute the errors on the blue points.

Report the average error on all points

k-Fold Cross Validation



Mean squared error: 2.05

Mean squared error: 1.11

Mean squared error: 2.93

What Kind of Cross Validation?

	Downside	Upside
Validation-set	Potentially Unreliable	Cheap
Leave-one-out	Expensive	Doesn't waste data
10-fold	Wastes 10% of the data. 10 times more expensive than validation-set	Only wastes 10%. Only 10 times more expensive instead of N times.
3-fold	Wastes more than 10-fold and more expensive than validation-set	Still better than validation-set
N-fold	Identical to Leave-one-out	

No ideal solution, but workable ones.