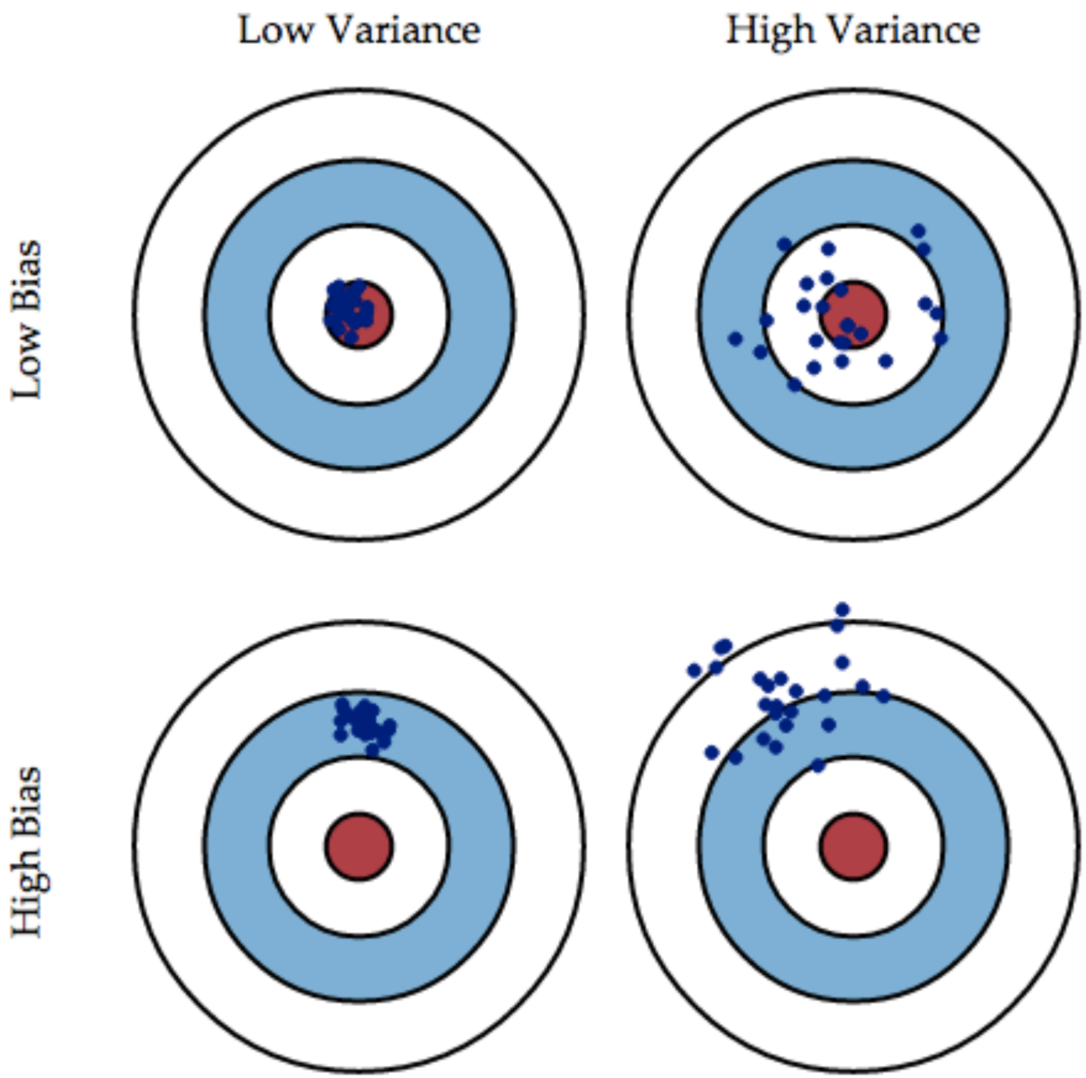
Variance-Bias tradeoff

Prediction error

Prediction error is driven by three factors:

(1) Variance
(2) Bias
(3) Noise

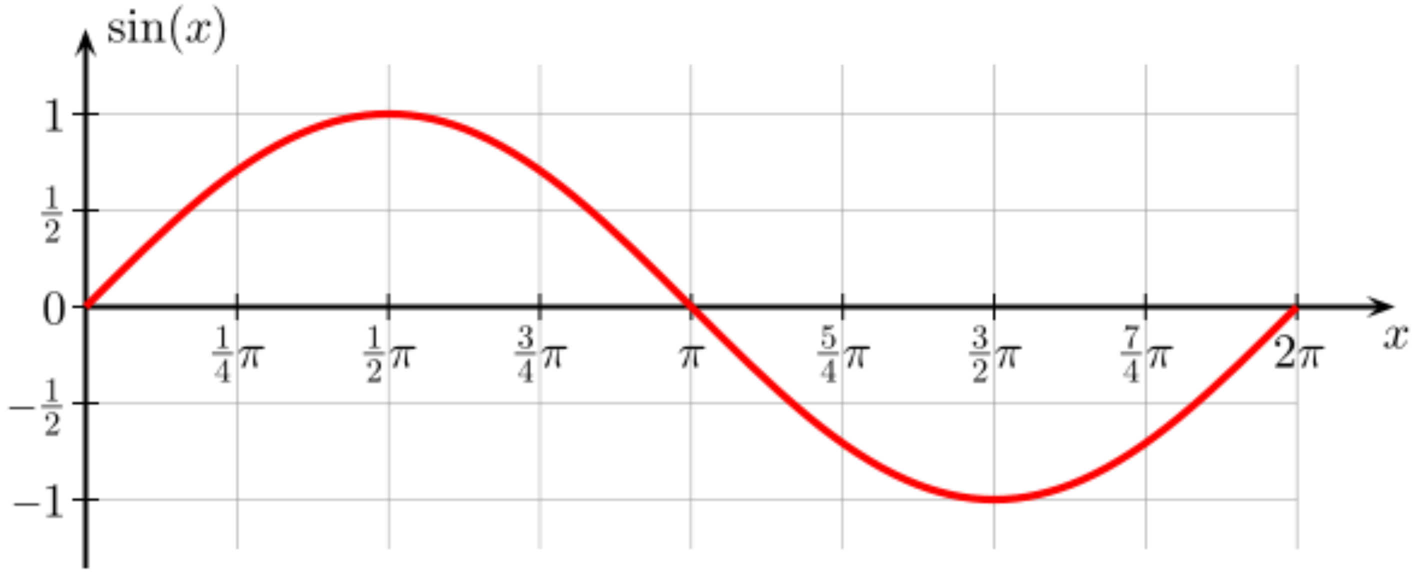
Variance-Bias Tradeoff



From: http://scott.fortmann-roe.com/docs/BiasVariance.html

Variance-Bias Tradeoff – another look

Suppose you are trying to learn the sine function:



Suppose also that your dataset consists of only two points.

We'll try two different models:

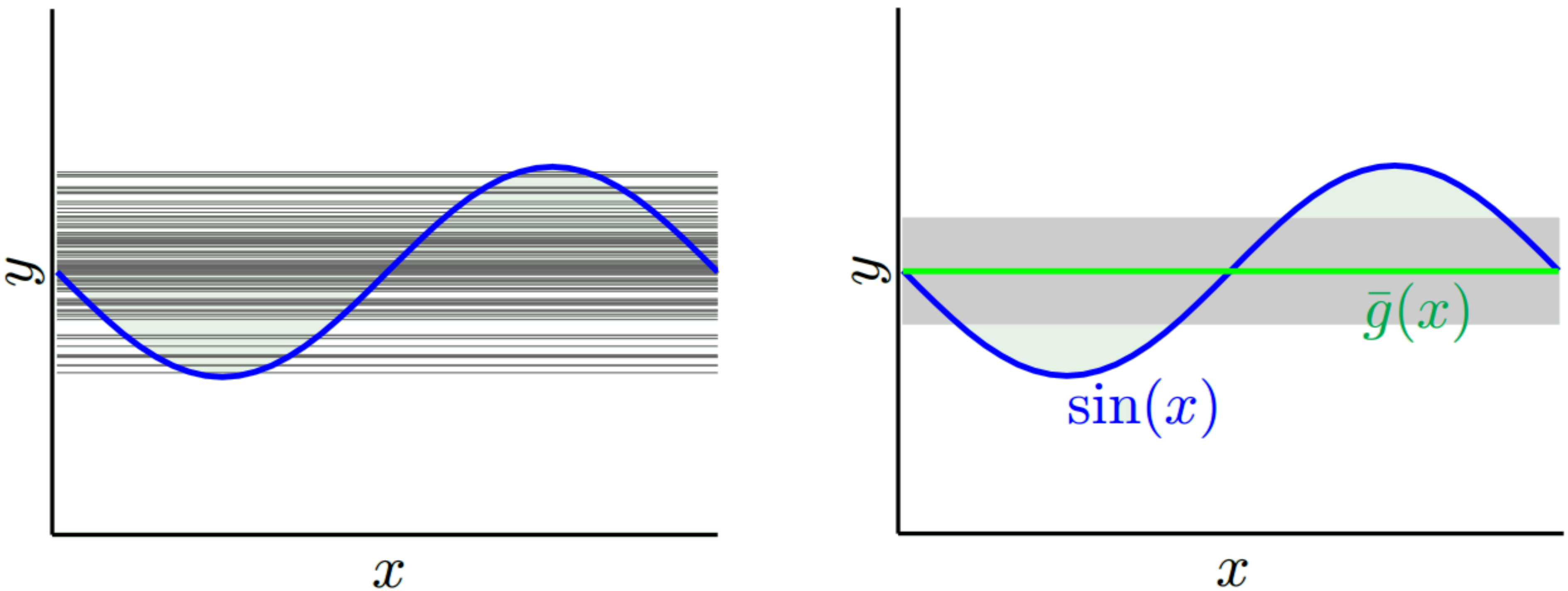
Constant: $H_0(x) = b$

Linear: $H_1(x) = ax + b$

From: http://work.caltech.edu/lectures.html#lectures

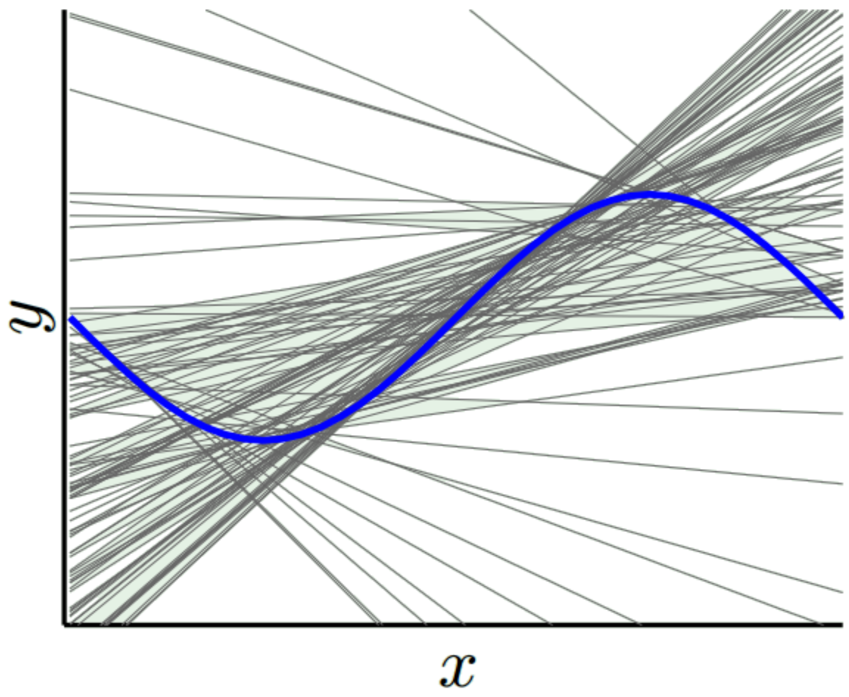
Constant: $H_0(x) = b$

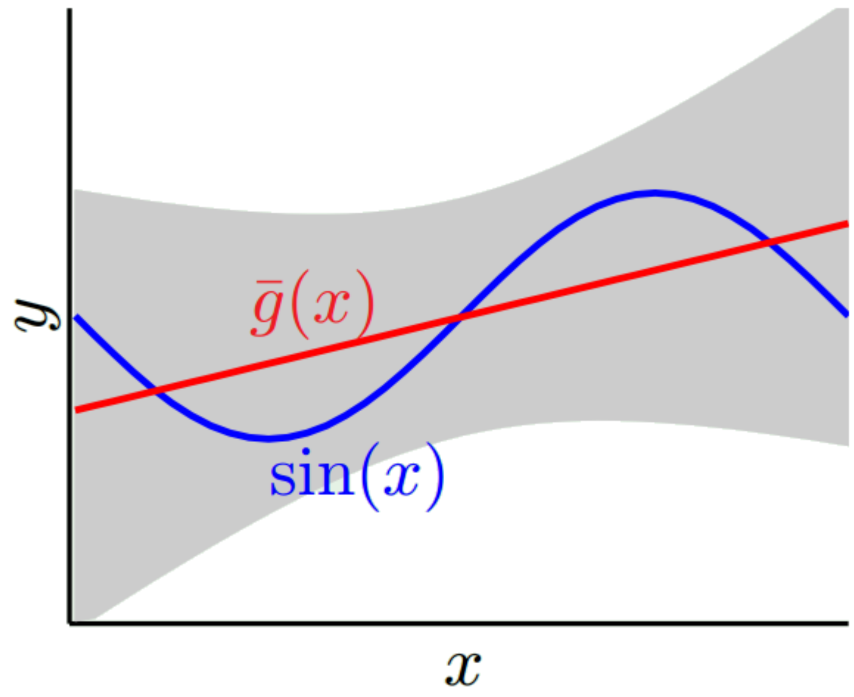
We use with many different training sets (i.e. we repeatedly select 2 data points and perform the learning on them), we obtain (left graph represents all the learnt models, right graph represent their mean g and their variance (grey area)):



Linear: $H_1(x) = ax + b$

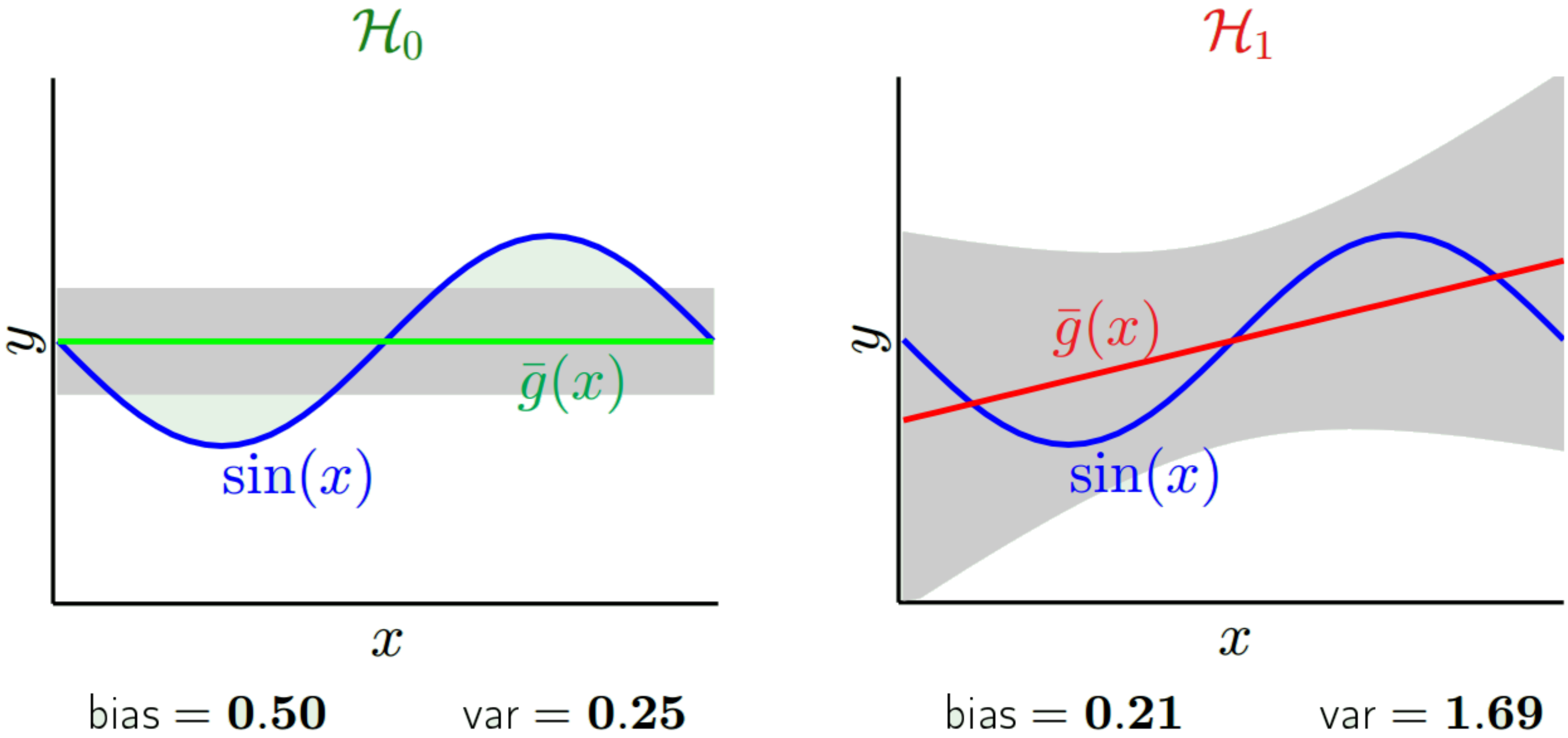
And we do the same for the linear model as well:



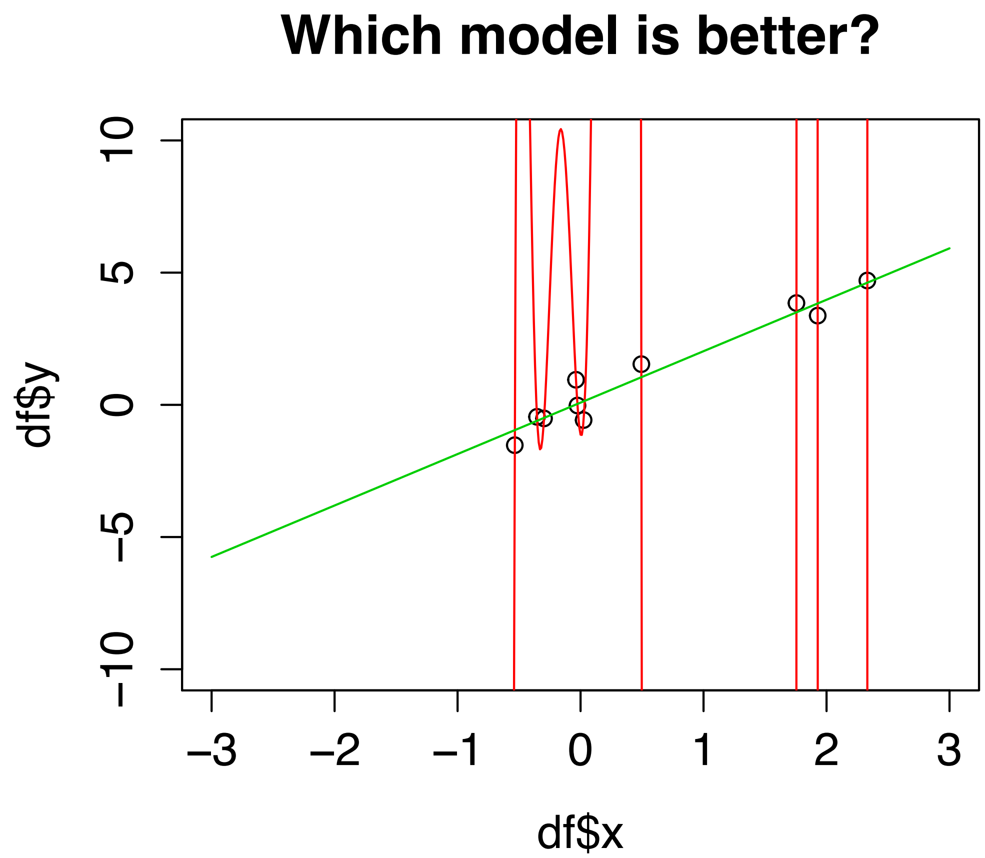


Compare

H_0 yields simpler models than H_1, hence a lower variance when we consider all the models learnt with H_0, but the best model g (in red on the graph) learnt with H_1 is better than the best model learnt g with H_0, hence a lower bias with H_1:



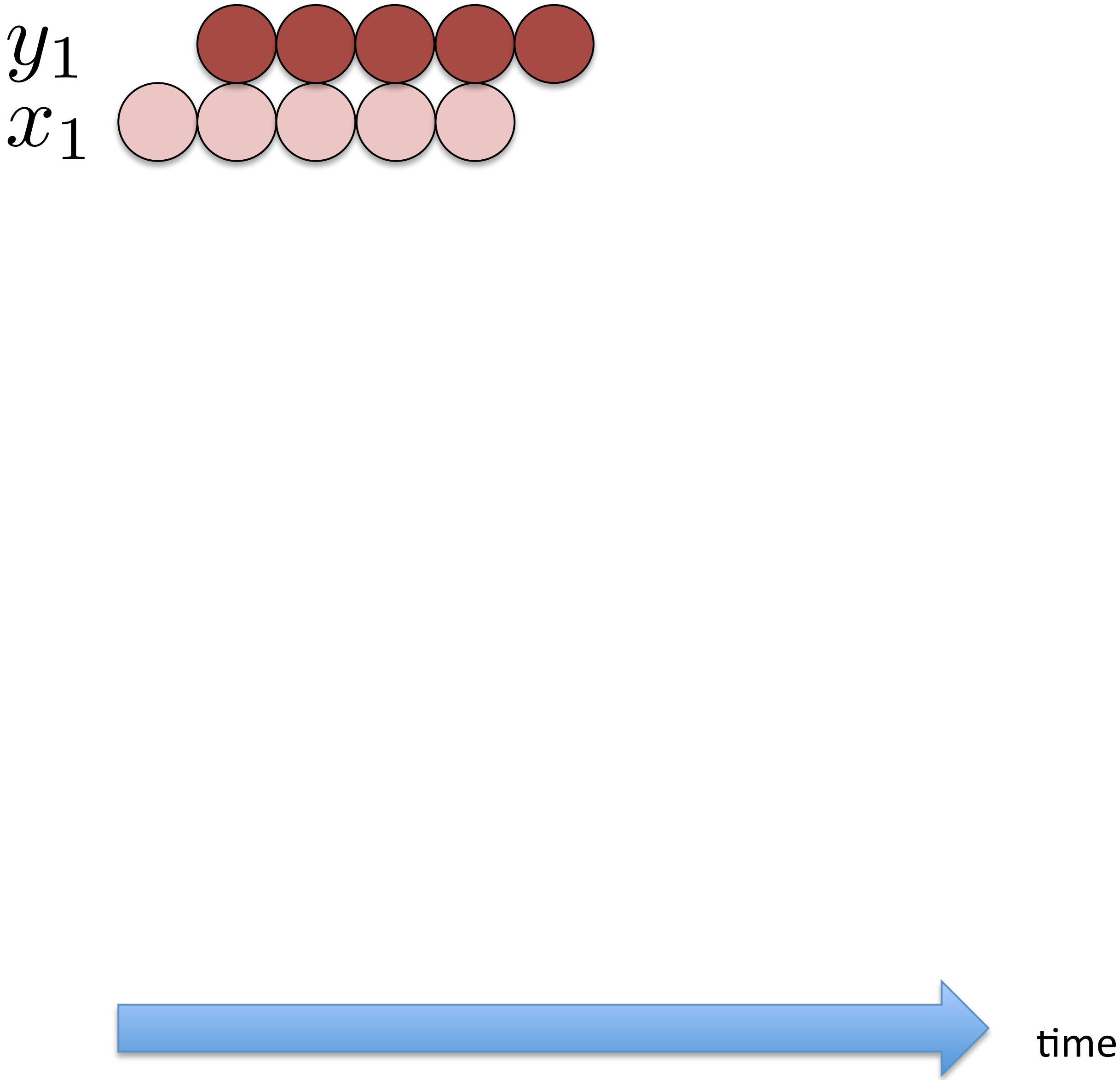
Red: 9^{th} order polynomial
Green: linear regression of x on y

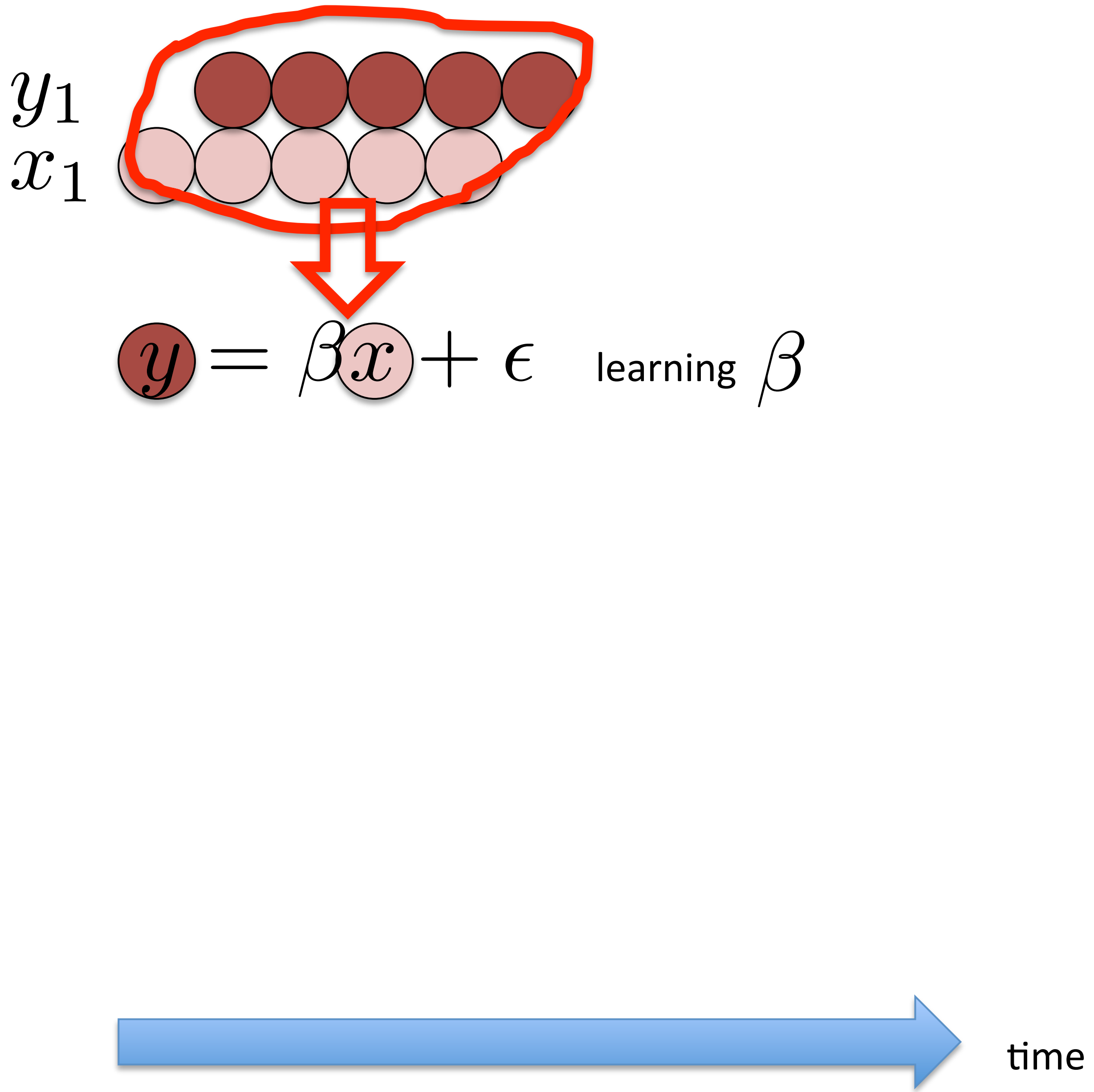


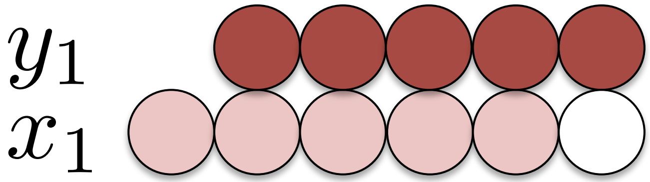
Rolling window analysis

Rolling Window Analysis

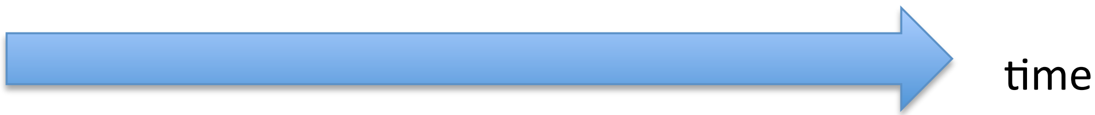
- For predictive analysis, using multiple regression, there are two factors one must consider
- Which factors to include in a model
- How much data one should include in the model

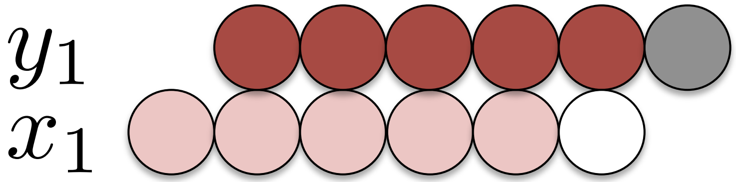

y_1
x_1
time


y_1
x_1
$y = \beta x + \epsilon$
learning β
time



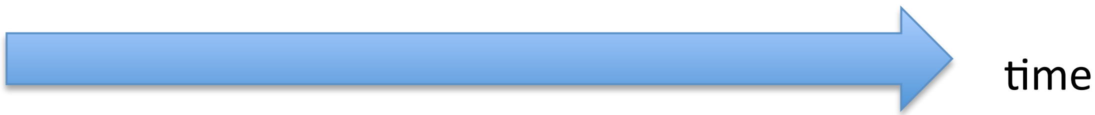
$y = \beta x + \epsilon$ learning β

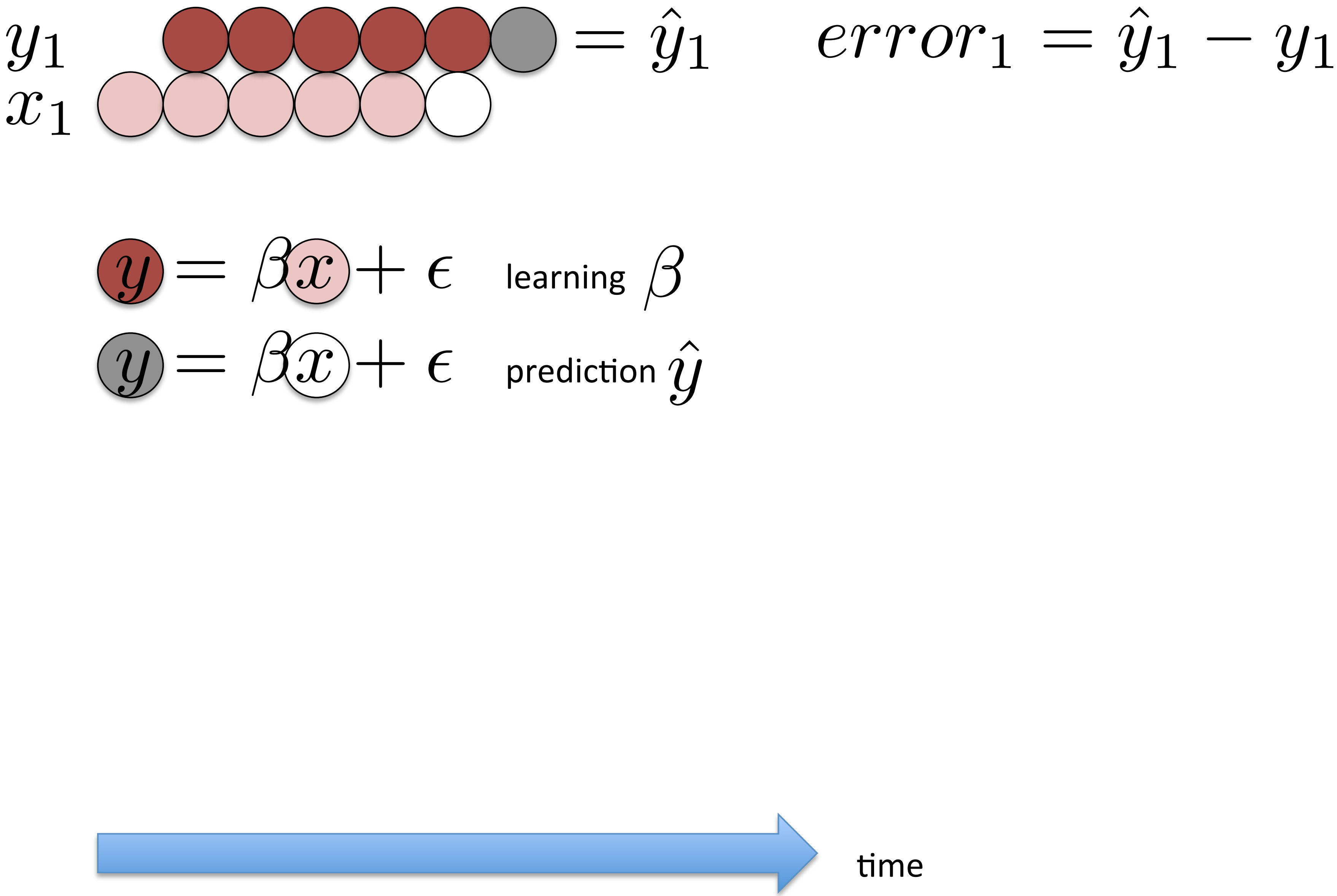




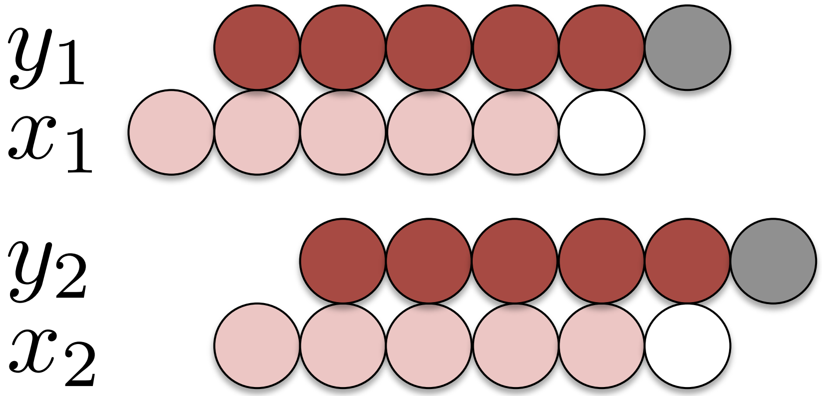
$$y = \beta x + \epsilon$$ learning β

$$y = \beta x + \epsilon$$ prediction $\hat{y}$



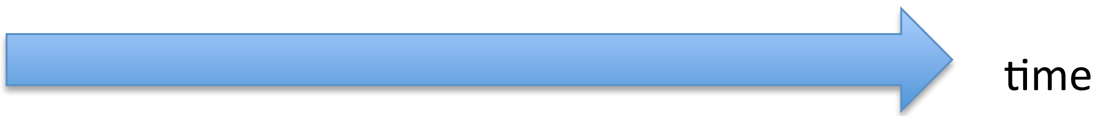


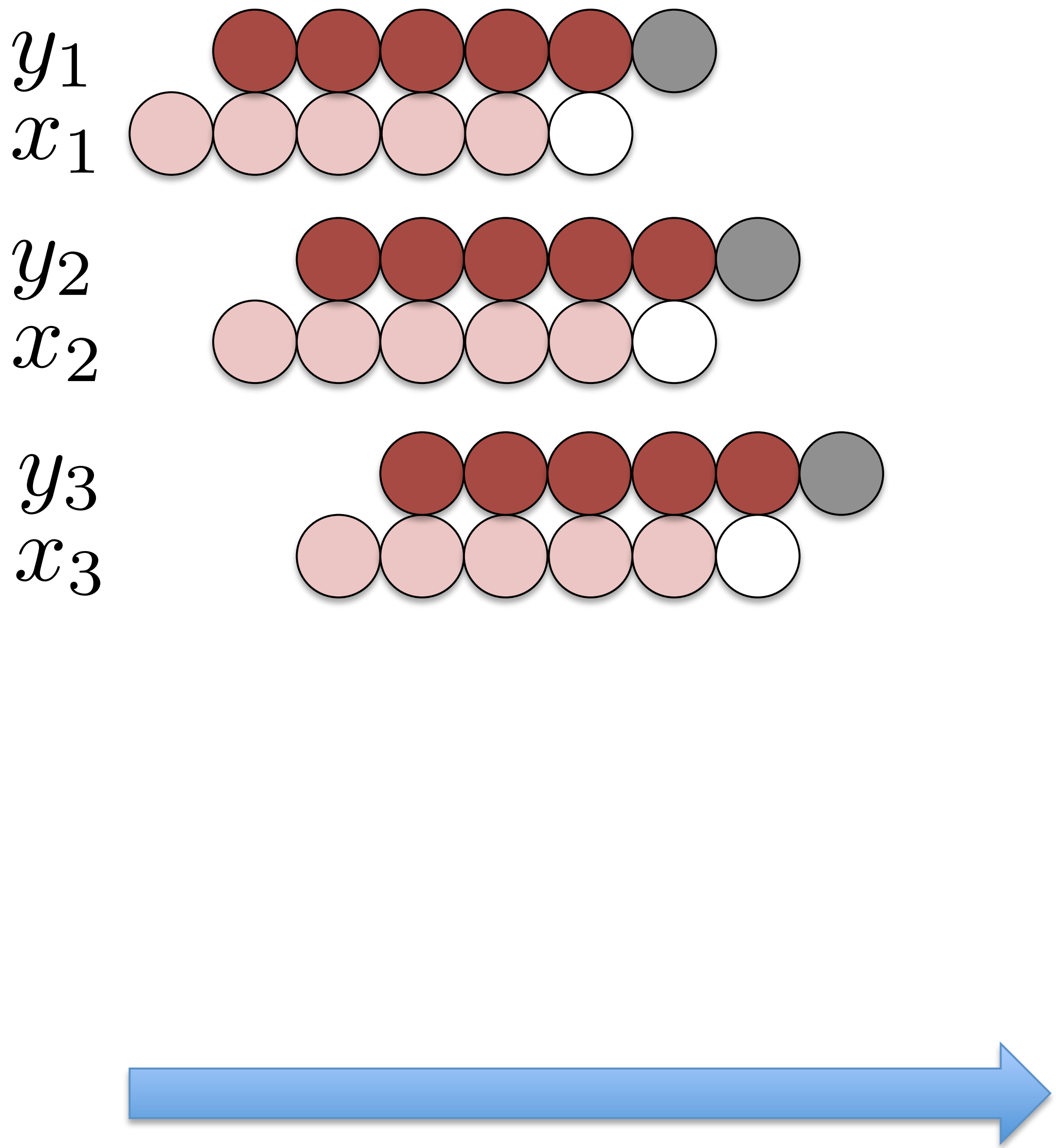
y_1
x_1
$= \hat{y}_1$
$error_1 = \hat{y}_1 - y_1$
$y = \beta x + \epsilon$
learning β
$y = \beta x + \epsilon$
prediction $\hat{y}$
time



$$error_1 = \hat{y}_1 - y_1$$

$$error_2 = \hat{y}_2 - y_2$$



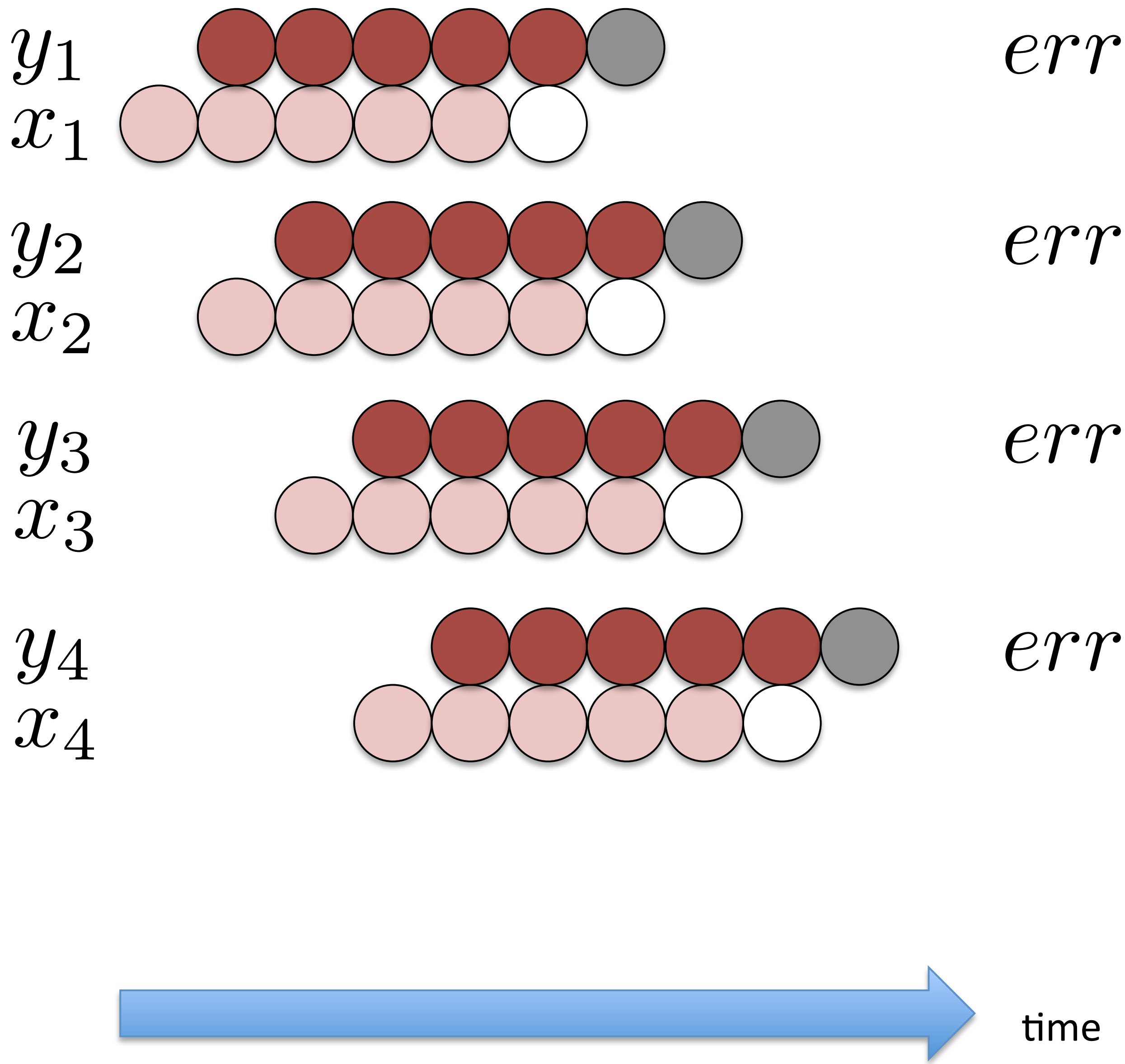


time

$$error_1 = \hat{y}_1 - y_1$$

$$error_2 = \hat{y}_2 - y_2$$

$$error_3 = \hat{y}_3 - y_3$$

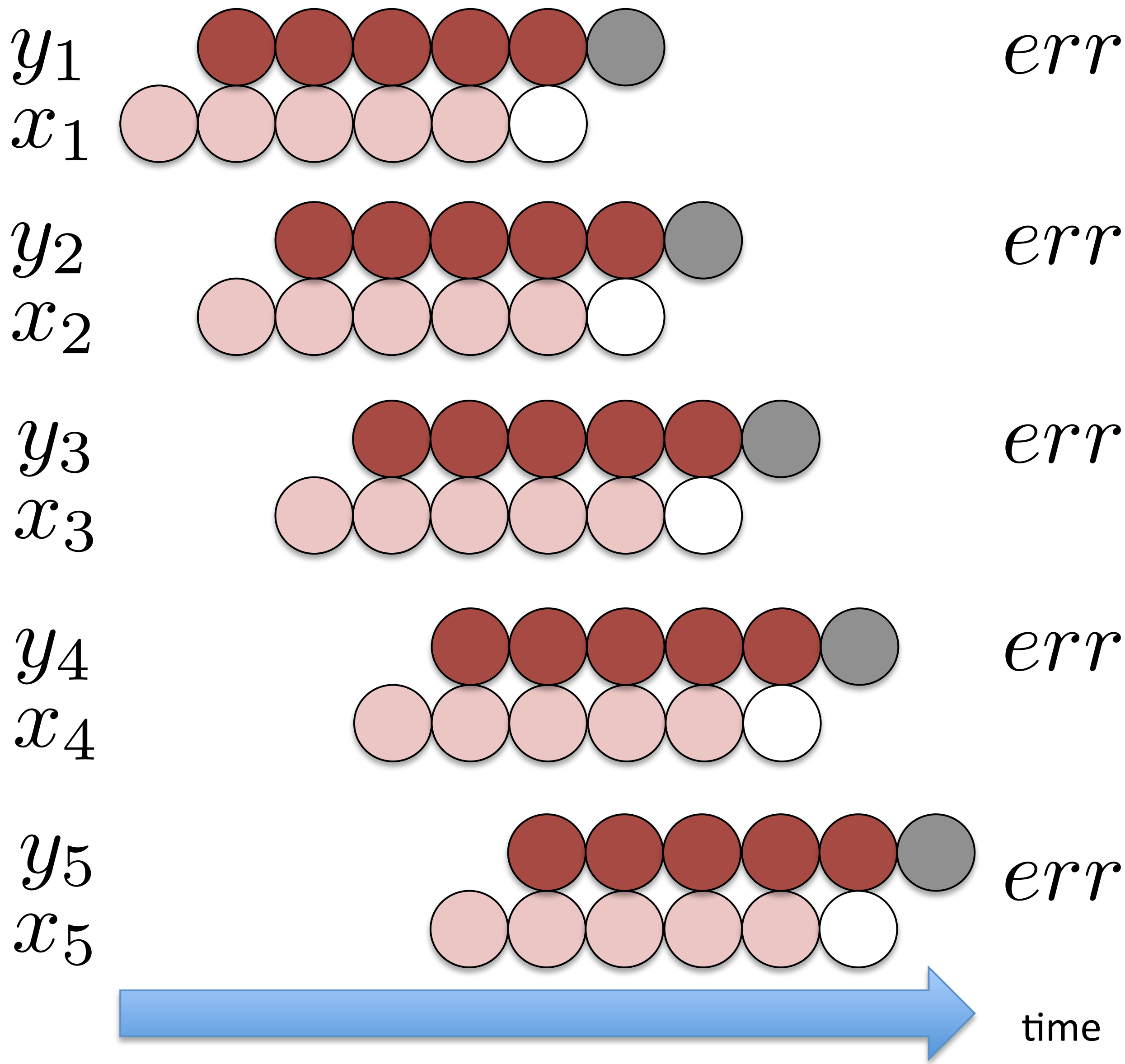


$$error_1 = \hat{y}_1 - y_1$$

$$error_2 = \hat{y}_2 - y_2$$

$$error_3 = \hat{y}_3 - y_3$$

$$error_4 = \hat{y}_4 - y_4$$



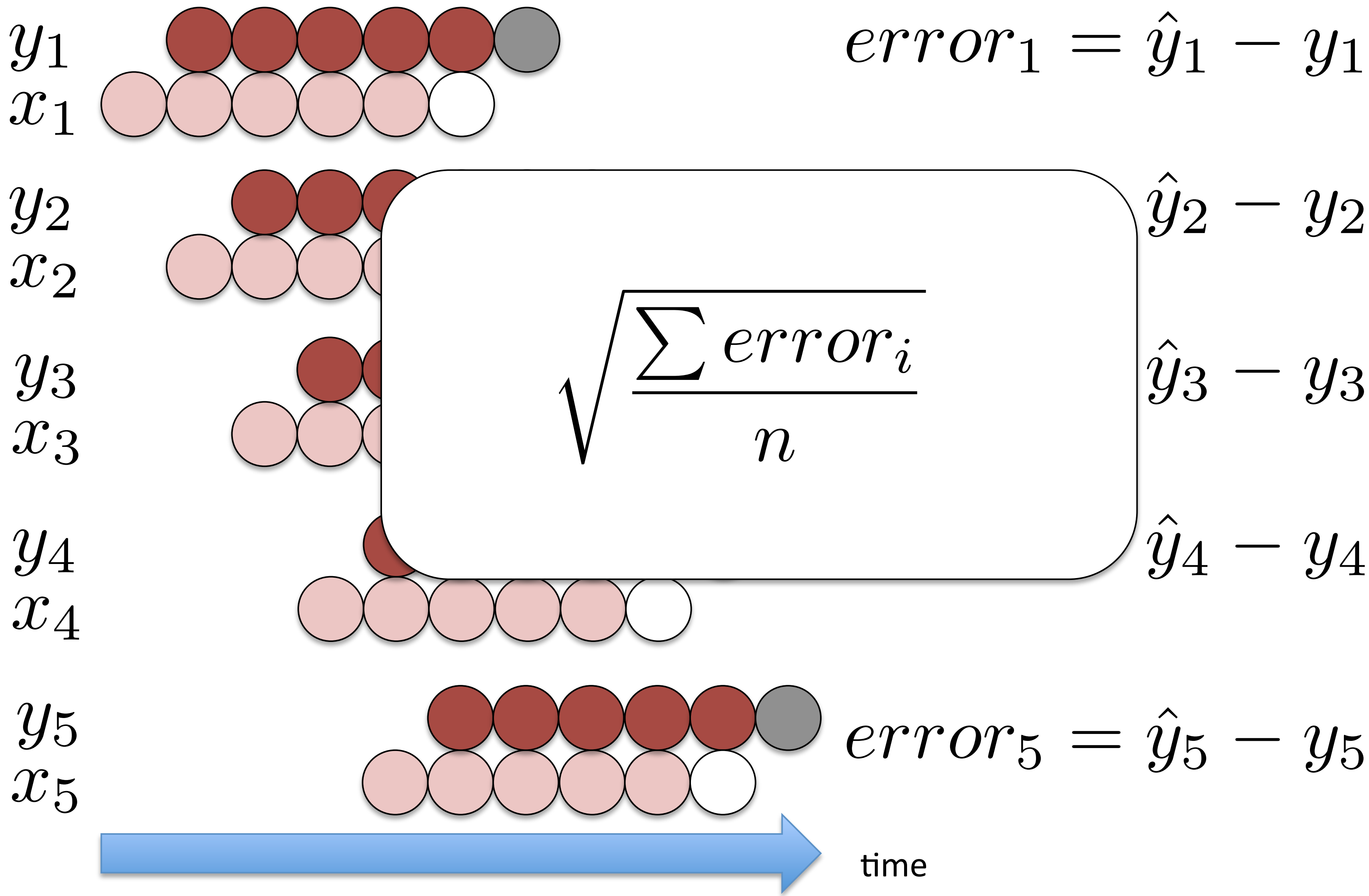
$$error_1 = \hat{y}_1 - y_1$$

$$error_2 = \hat{y}_2 - y_2$$

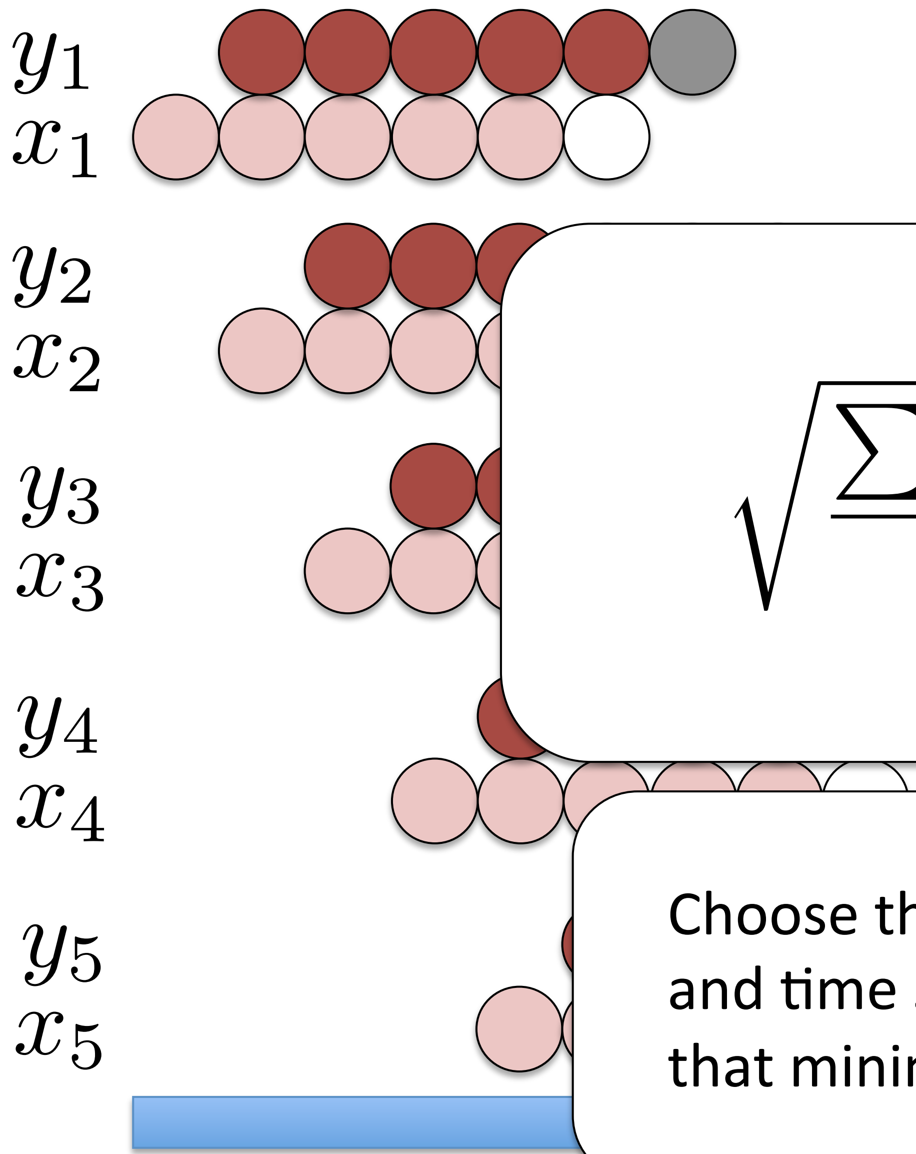
$$error_3 = \hat{y}_3 - y_3$$

$$error_4 = \hat{y}_4 - y_4$$

$$error_5 = \hat{y}_5 - y_5$$



y_1
x_1
y_2
x_2
y_3
x_3
y_4
x_4
y_5
x_5
$error_1 = \hat{y}_1 - y_1$
$\hat{y}_2 - y_2$
$\hat{y}_3 - y_3$
$\hat{y}_4 - y_4$
$error_5 = \hat{y}_5 - y_5$
$\sqrt{\frac{\sum error_i}{n}}$
time



$$error_1 = \hat{y}_1 - y_1$$

$$\hat{y}_2 - y_2$$

$$\hat{y}_3 - y_3$$

$$\hat{y}_4 - y_4$$

$$\sqrt{\frac{\sum error_i}{n}}$$

Choose the combination of factors and time span (rolling window size) that minimizes the RMSE

Tasks & Notes

- Work on HW5, make sure you get a chance to discuss your approach with Joe or Ken during the session today
- Joe & Ken (last) OH: Monday 4:30 – 5:30pm
- Prof. C-R OH: Monday 3:30-4:30pm