Advanced Python for Neuroscientists

Lecture 3: Regression, Classification (Decoding)

2022/07/05

Git Command

Git(hub)

- Git - software for tracking changes in any set of files, usually used for coordinating work among programmers collaboratively developing source code during software development. `conda install numpy git`

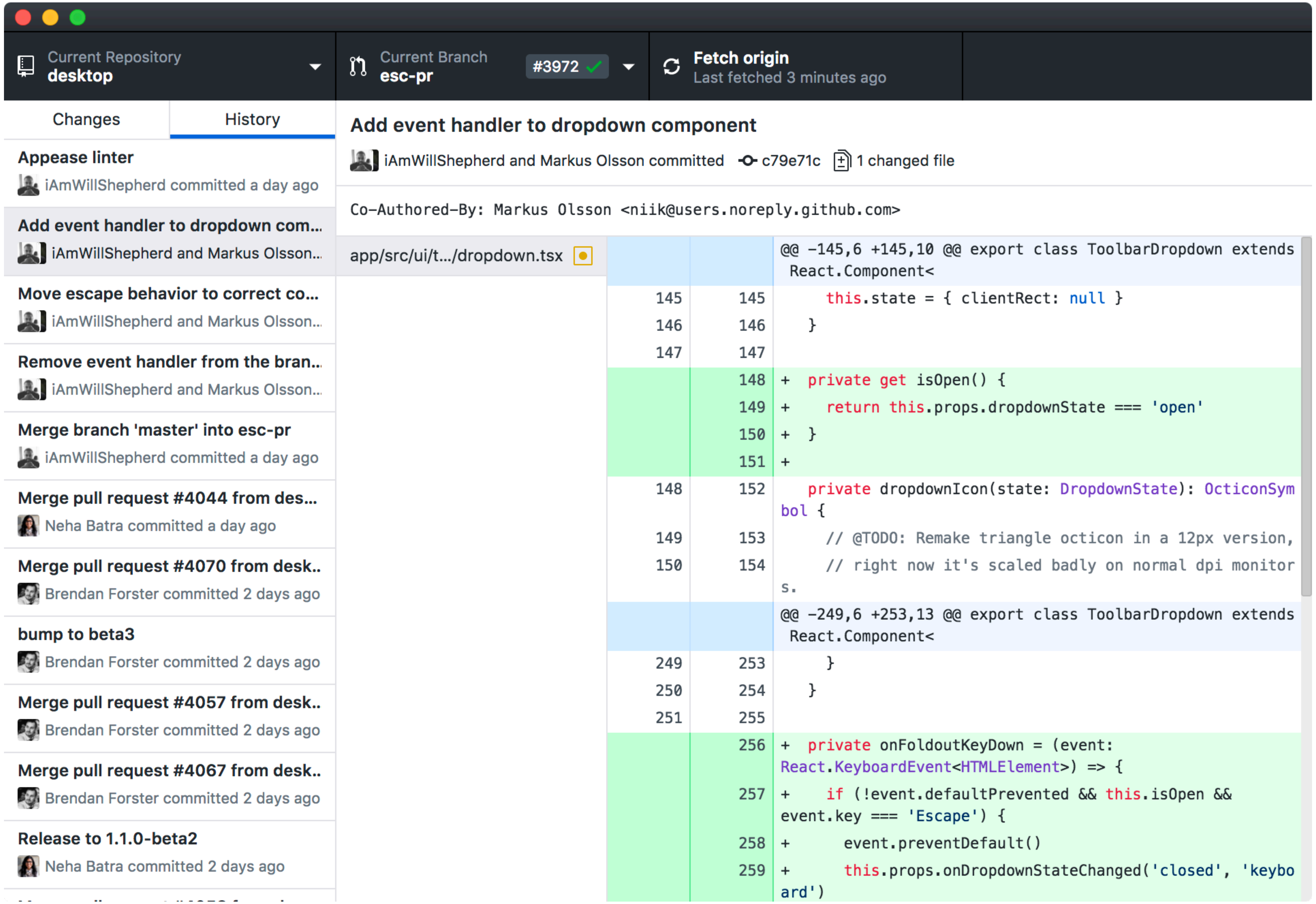
- GitHub, Inc. - a provider of Internet hosting for software development and version control using Git

- Versions with new changes - commits

- Upload - push; download - pull

User interface

- Github desktop



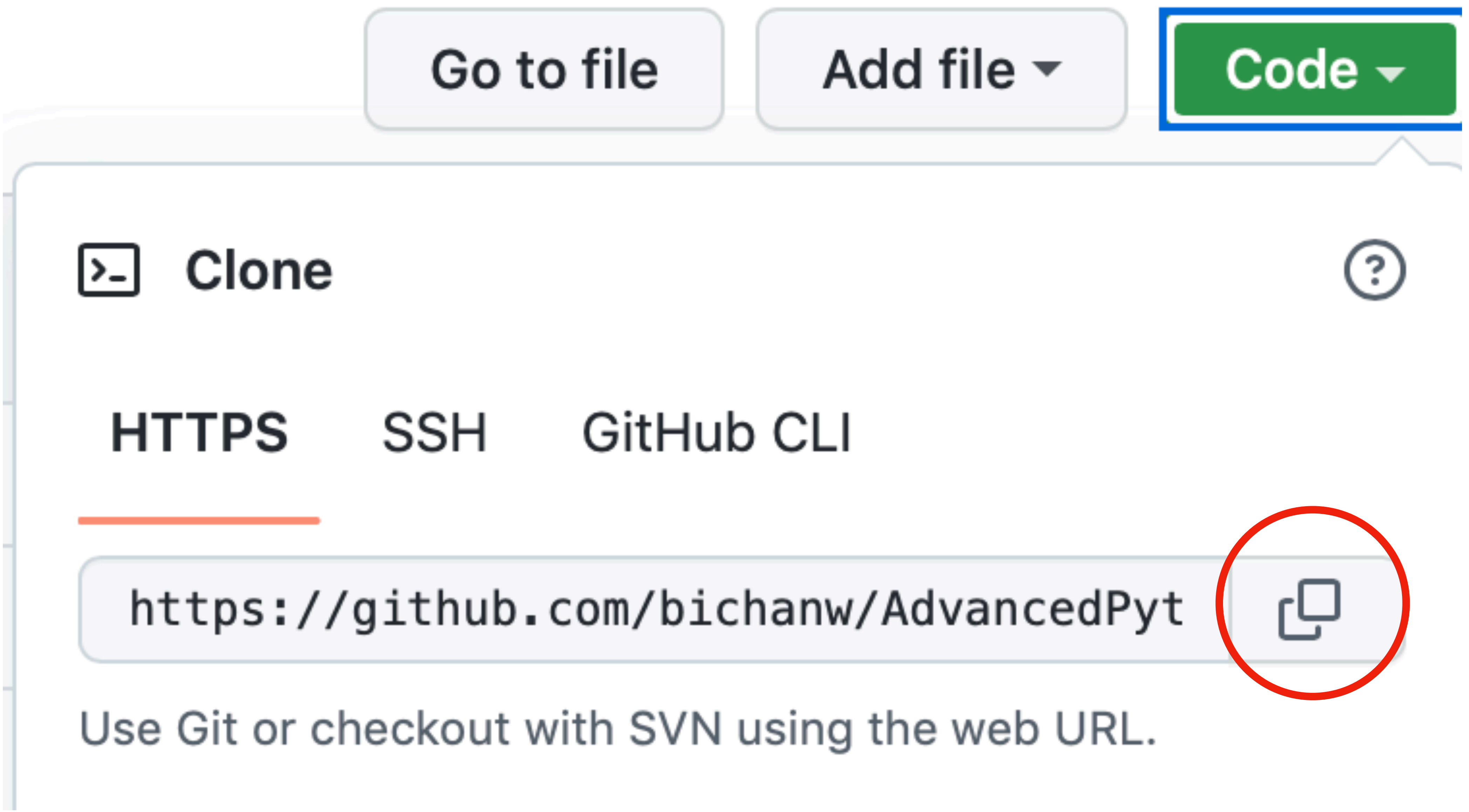
User interface

- Github desktop
- Sublime merge

User interface

Command lines

- First download: git clone



User interface

Command lines

- First download: git clone + repo location
- Update new changes to local: git pull
- Update local changes to GitHub:
  - Add local changes to a new commit: git add -A / file name
  - Specify what changes are made: git commit -m “name the changes”
  - Upload changes: git push

User interface

Command lines

- Cancel local changes: git checkout filename
- Hard reset:
  - git fetch --all
  - git branch backup-master
  - git reset --hard origin/master
- Put them all in a bash script (example)

Object-oriented Programming

Object-oriented programming

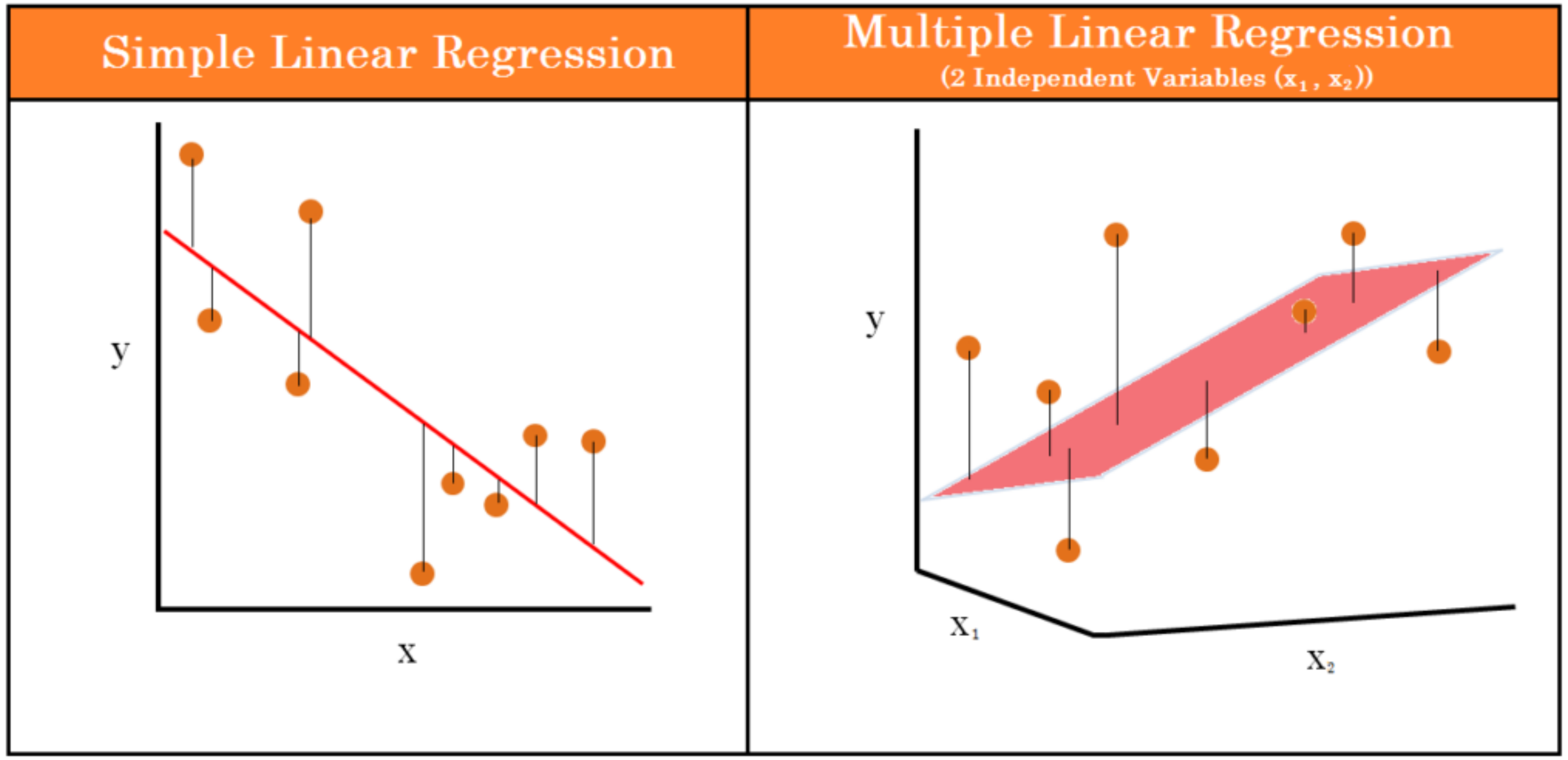
- An object with packs of values + codes
  - Dictionary (struct) + functions
- Easier for organizing and naming functions

- dog_name = dog()
- dog_name.bark(Y)

Regression

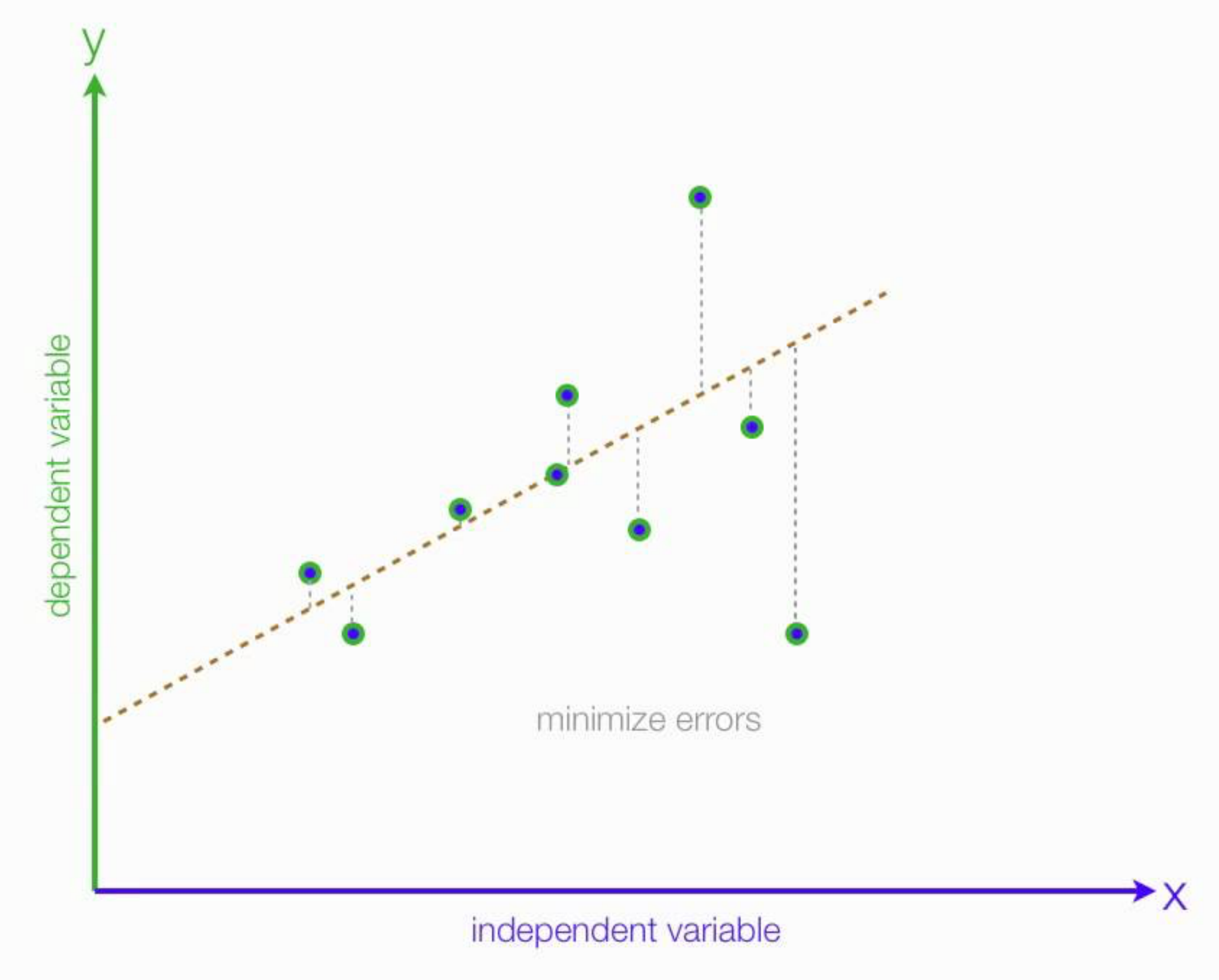
Linear Regression

- Predicting a real-valued output y
- given a vector of real valued inputs $x_1, ..., x_n$
- Assuming their relationship is linear: $y = w_1 x_1 + ... w_n x_n + w_0$



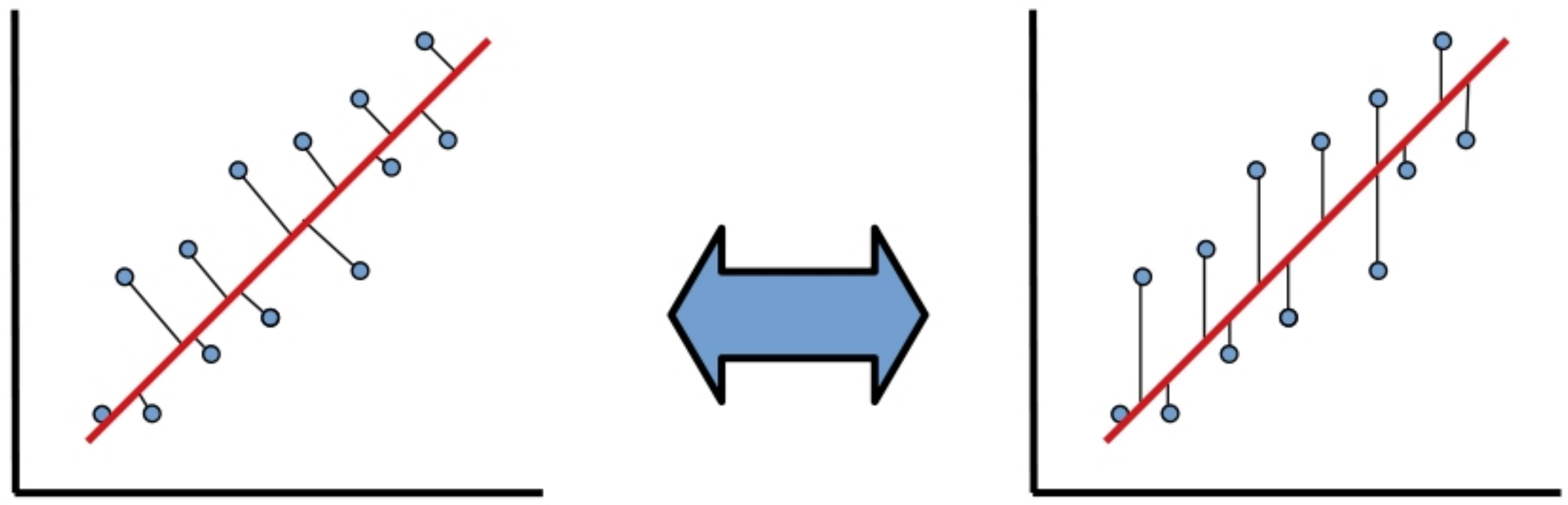
How to find w

- Find the line the best fits the data
- Predict y with x
- Minimize difference between y and $y_{\text{predicted}}$



PCA vs linear regression

- Distance to the principal axis vs. distance between $y, \hat{y}$
- Treat 2 variables as the same vs. input & output

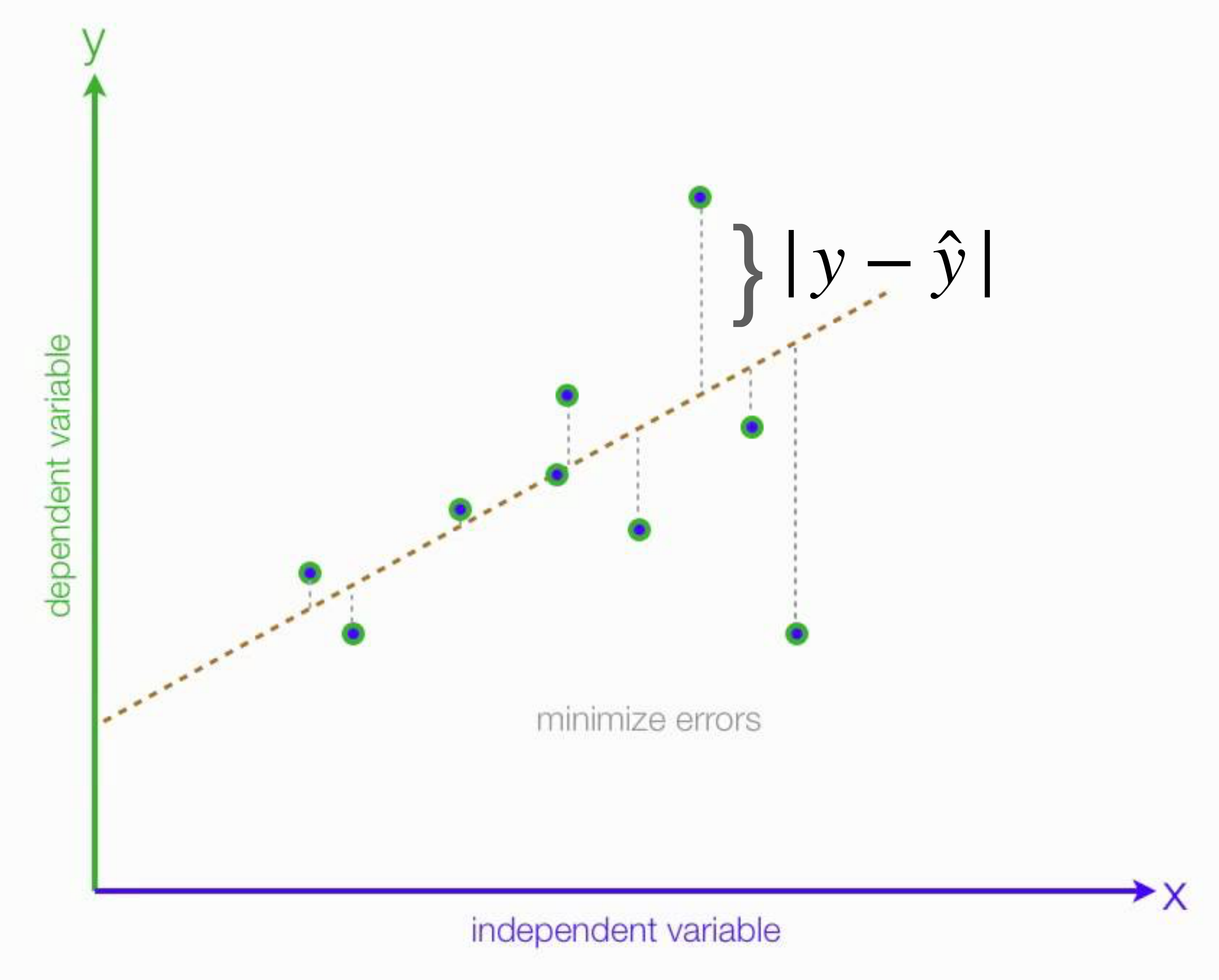


- estimation

- XB = Y
- XTX B = XT Y

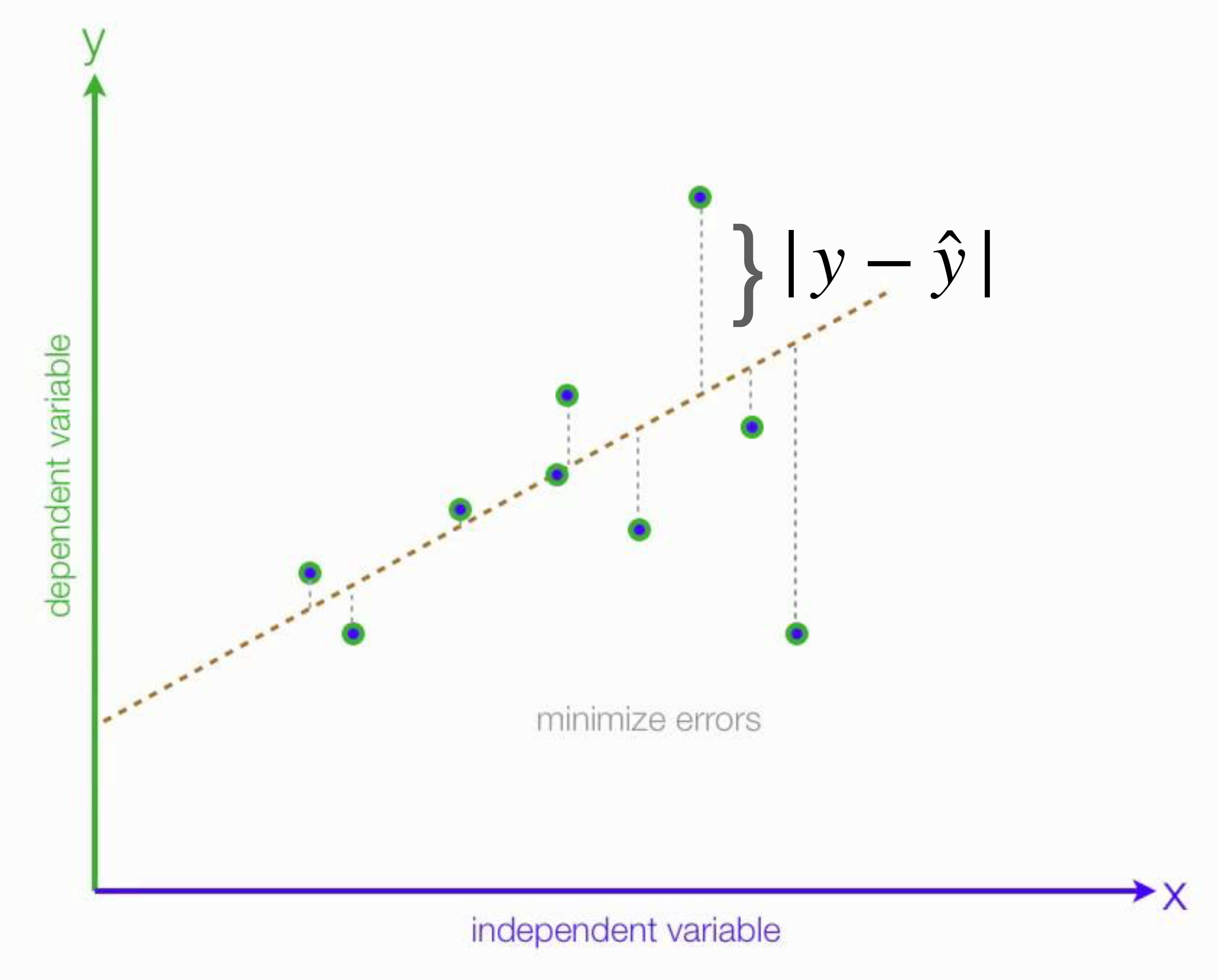
$$\mathrm{NLL}\left(\boldsymbol{w}, \sigma^2\right) = -\sum_{n=1}^{N} \log\left[\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}\left(y_n - \boldsymbol{w}^\top \boldsymbol{x}_n\right)^2\right)\right]$$

$$(X^TX)^{-1}X^TY$$



How good is w

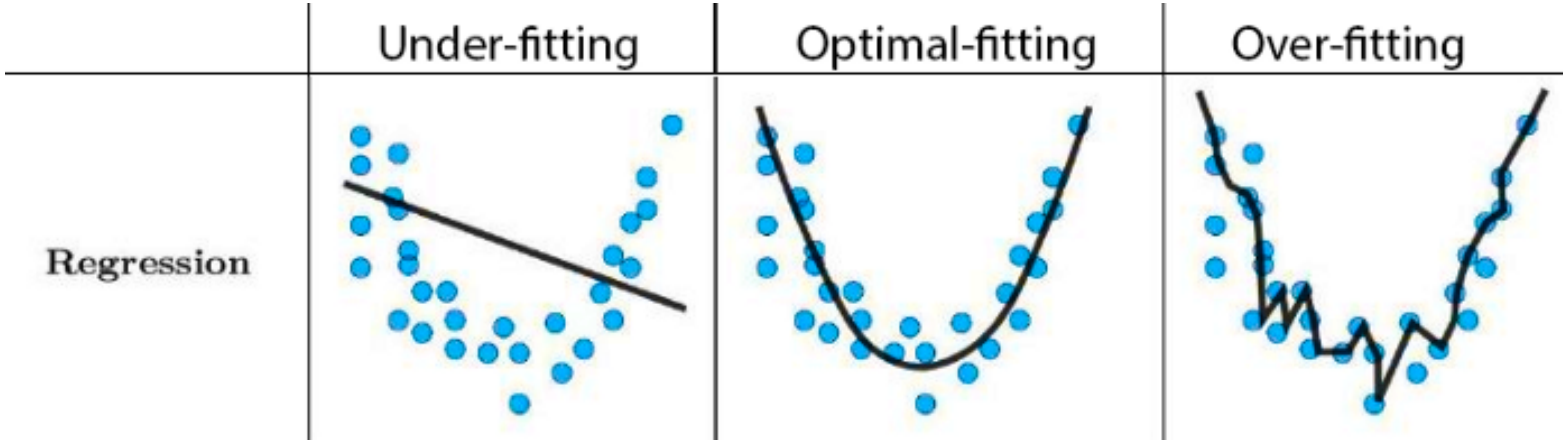
- $\sum (y - \hat{y})^2$ - least square estimation
- $1 - \dfrac{\sum (y - \hat{y})^2}{y^2}$
  - Coefficient of determination
  - Correlation coefficient (1d)
  - R^2
  - 1 - root mean squared error



Linear Regression Expansion

Add regularization

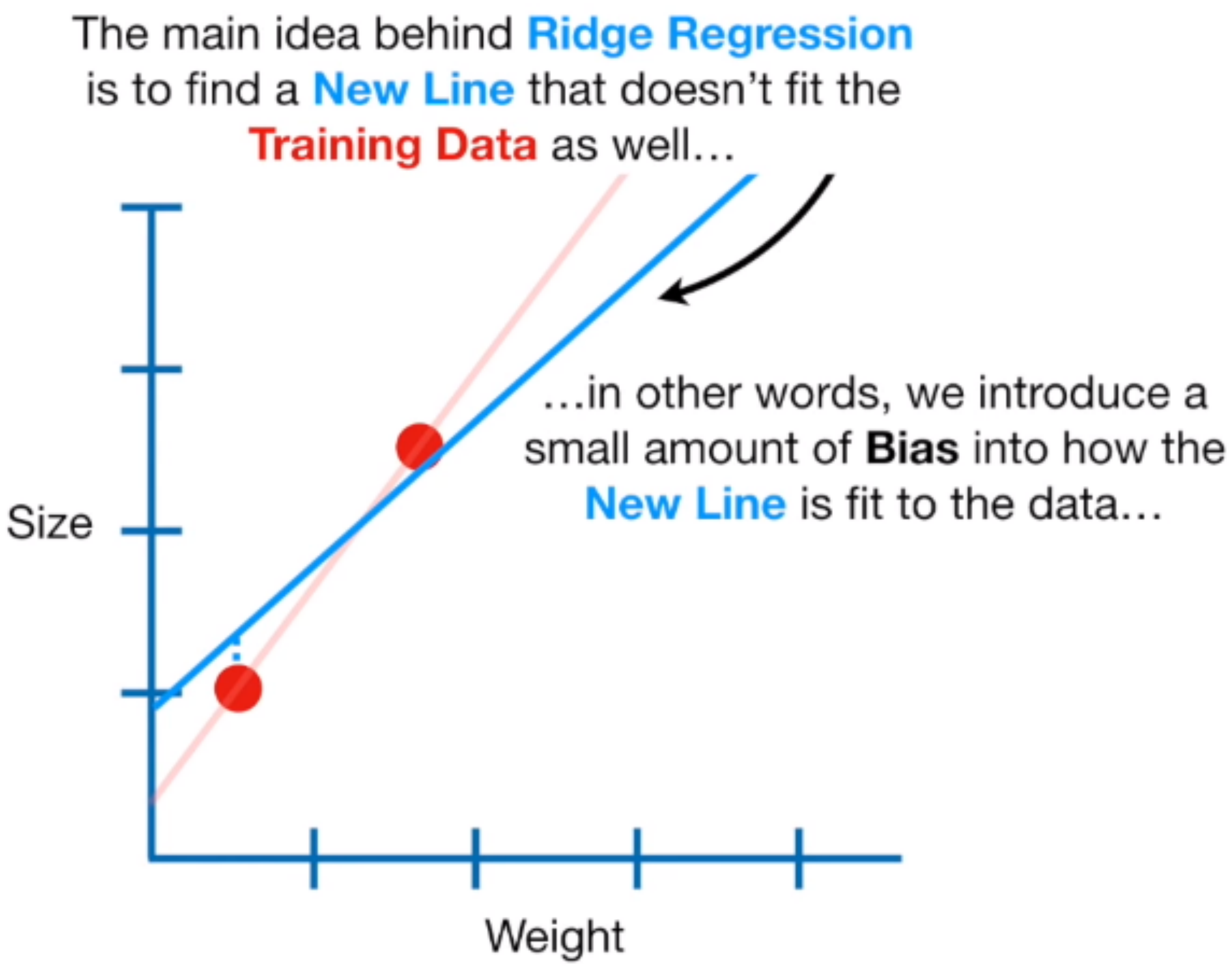
- Overfitting - the production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit to additional data or predict future observations reliably

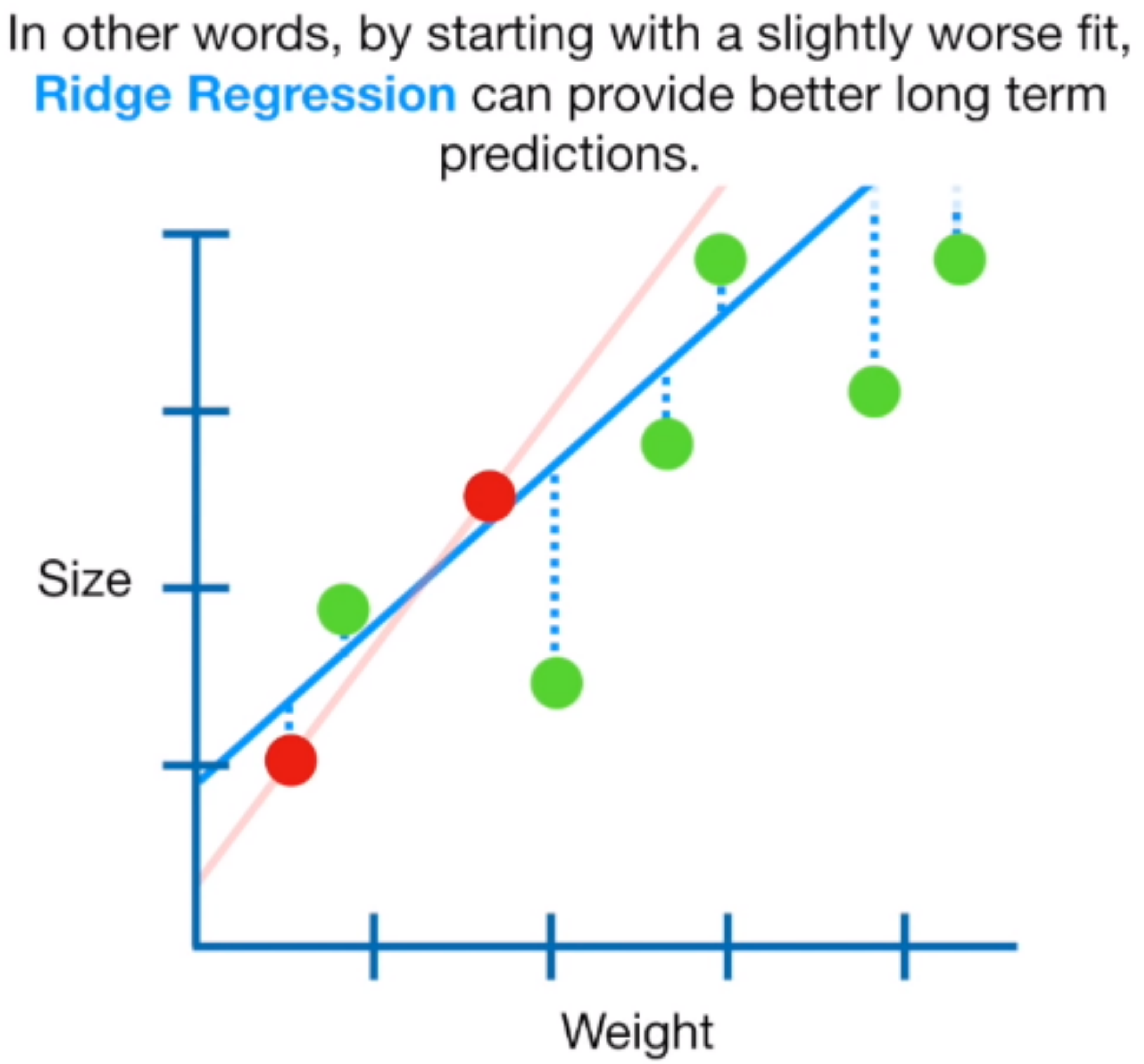


Linear Regression Expansion

Add regularization

- Overfitting





Linear Regression Expansion

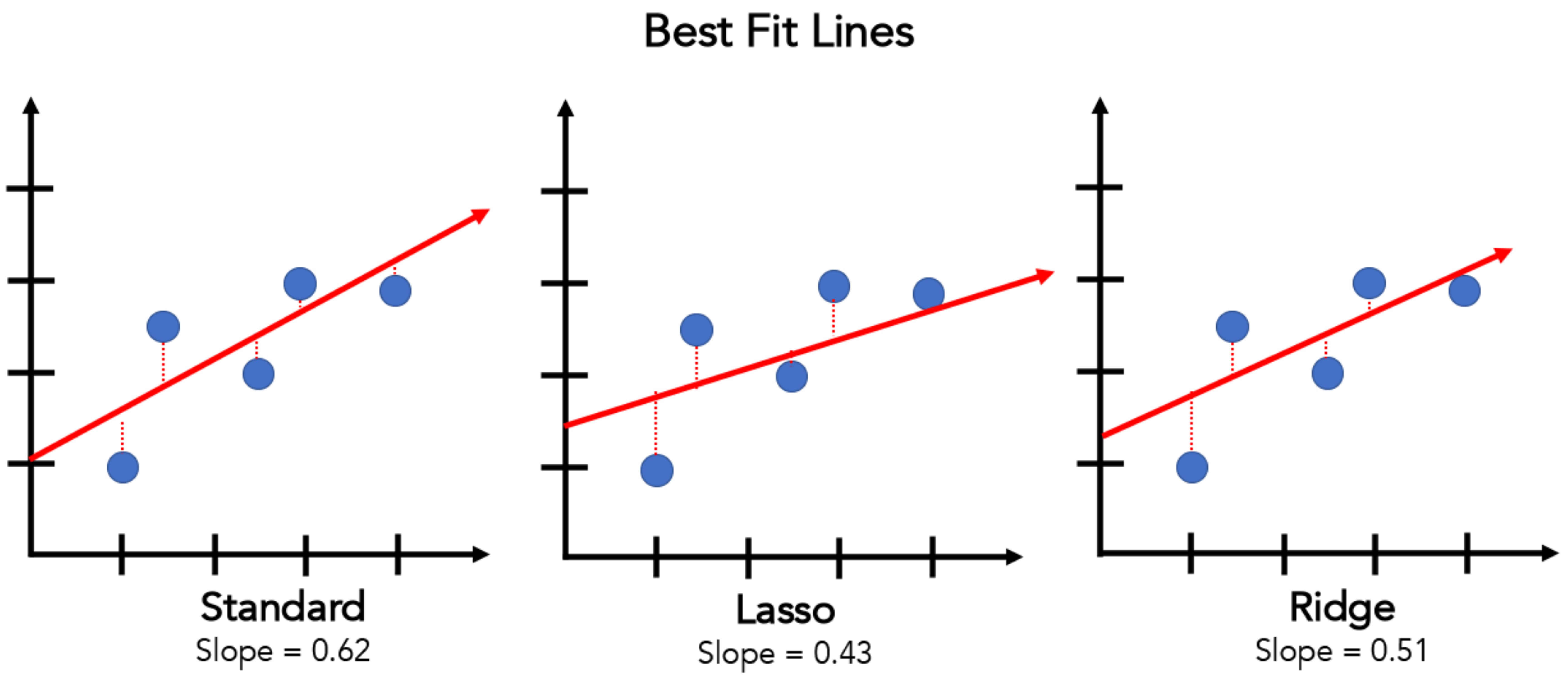
Add regularization

- Overfitting
- Regularization - make results “simpler”
  - Penalty term assumption: coefficients are not all large
  - Ridge (L2) regression: $\lambda \sum_i w_i^2 + \sum (y - \hat{y})^2$
  - Lasso (L1) regression: $\mathrm{RSE} + \lambda \sum_i |w_i|$

Linear Regression Expansion

Add regularization

- Overfitting



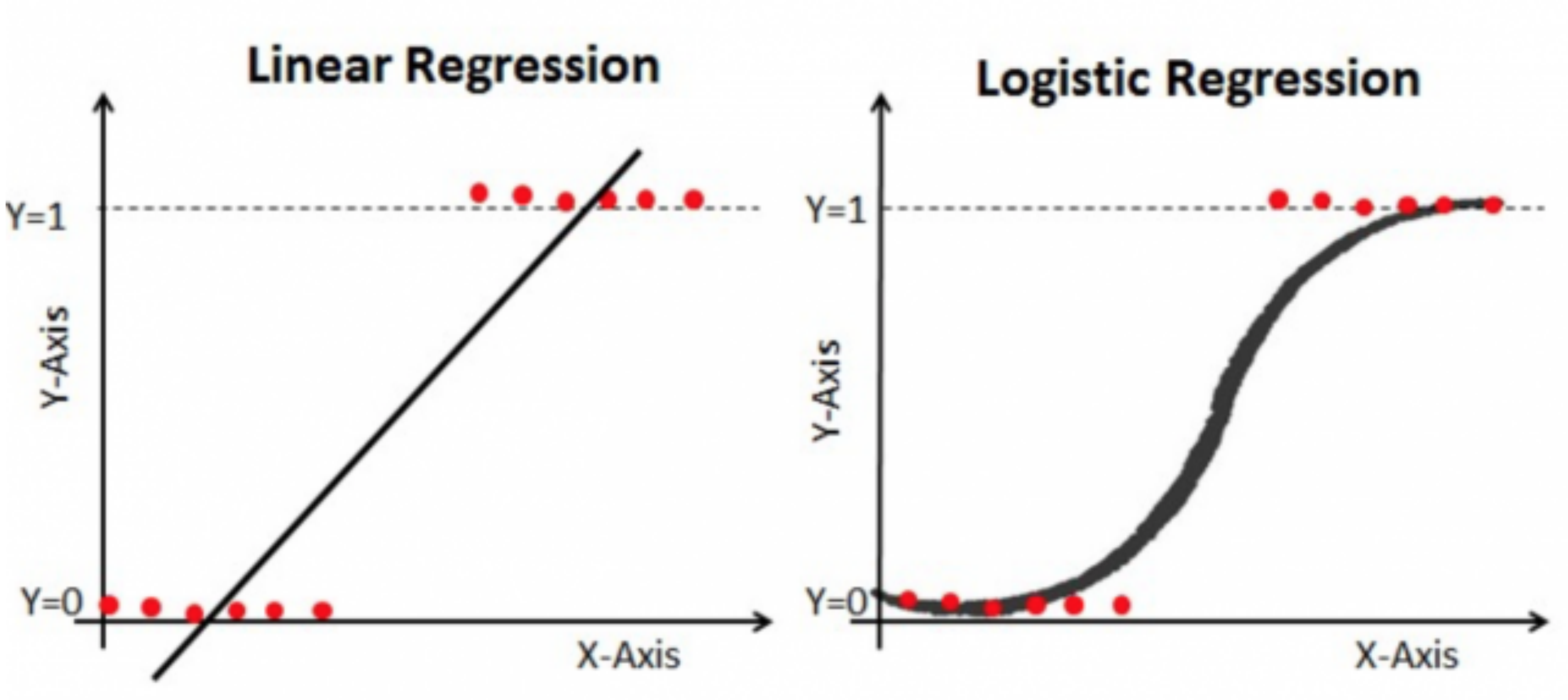
Linear Regression Expansion

Add nonlinear terms

- $y = w_1 x_1 + \ldots w_n x_n + w_0$
- What if true relationship is ~~slightly~~ more complicated: $y = x_1 + 3x_2^2$
- Simply calculate $x_3 = x_2^2$, run linear regression on x_1, x_2, x_3
- You need to have an idea what the nonlinear term is (e.g. $\sin x, x_1 x_2$)

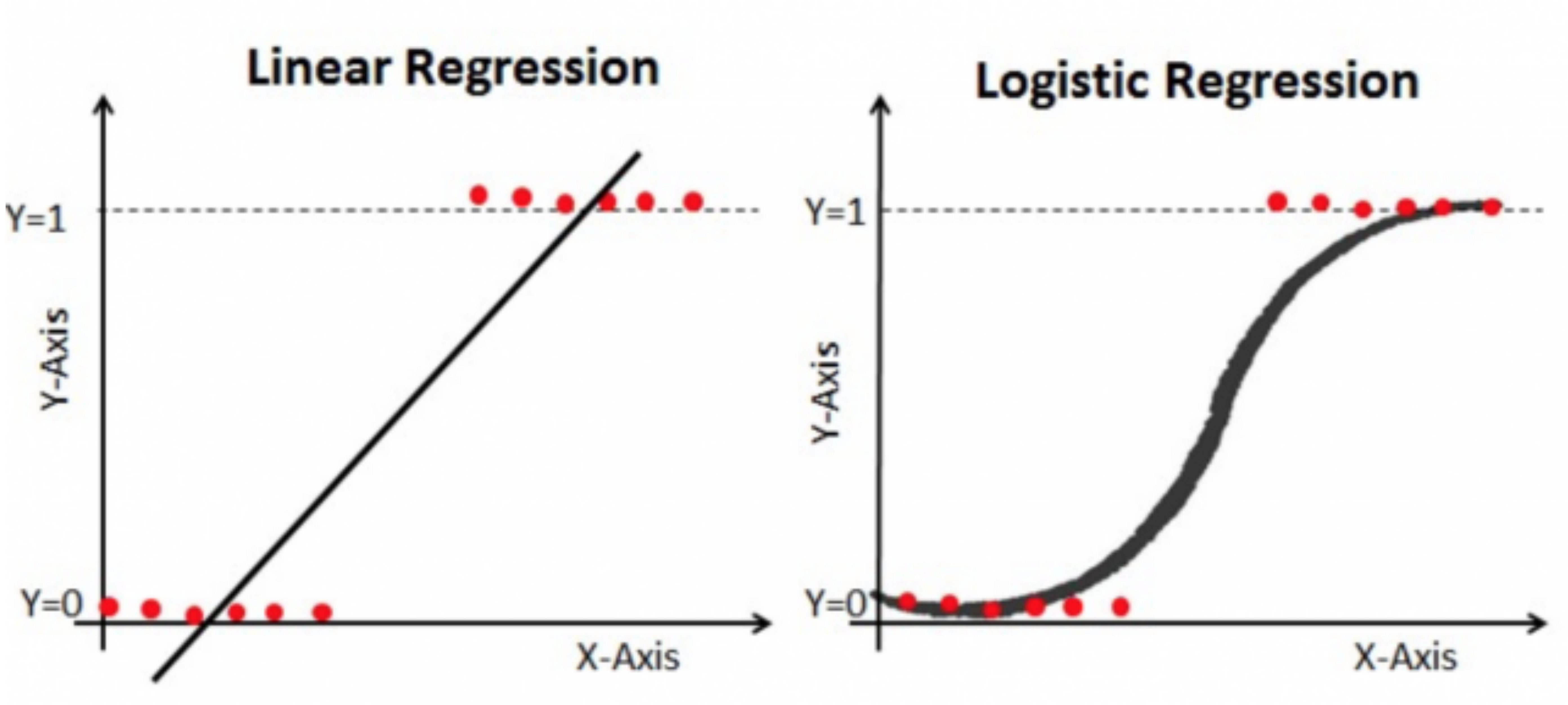
Logistic Regression

- Predicting a binary y



Logistic Regression

- Predicting a binary y
- Add a transformation to compress y - logistic function
- $y = \dfrac{1}{1 + e^{w_1 x_1 + w_2 x_2 \ldots w_0}}$



Logistic Regression

- Predicting a binary y
- Add a transformation to compress y - logistic function
- $y = \dfrac{1}{1 + e^{w_1x_1+w_2x_2...w_0}}$
- Commonly used in binary outcome tasks (selecting left / right)