Holdout Validation

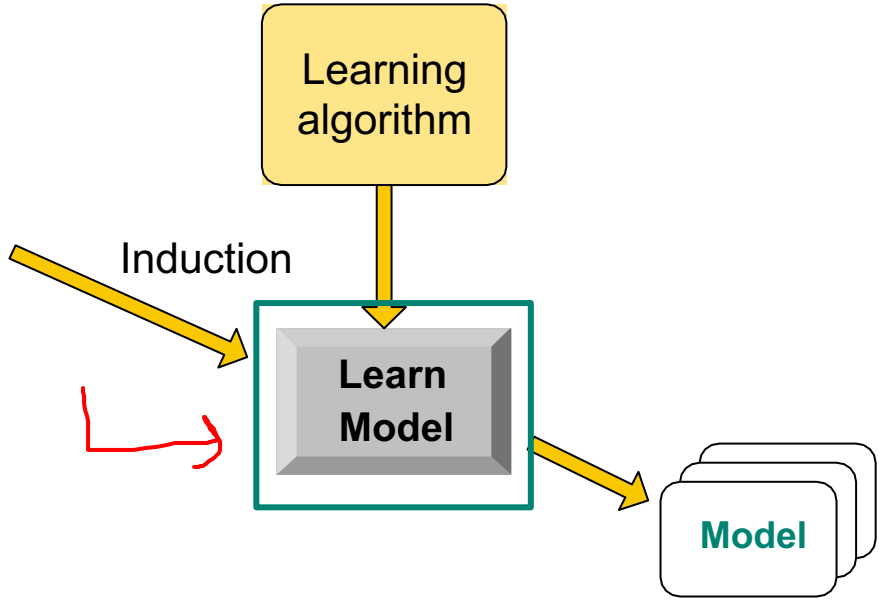
Lesson

AI Academy

Illustrating Classification Task

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Training Set



Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

Test Set

A learning system: basic cycle

1) Data: $D = \{d_1, d_2, ..., d_n\}$

2) Choose a learner:
Select a model with parameters (e.g. Decision Tree).

Parameter: A model value learned in training.

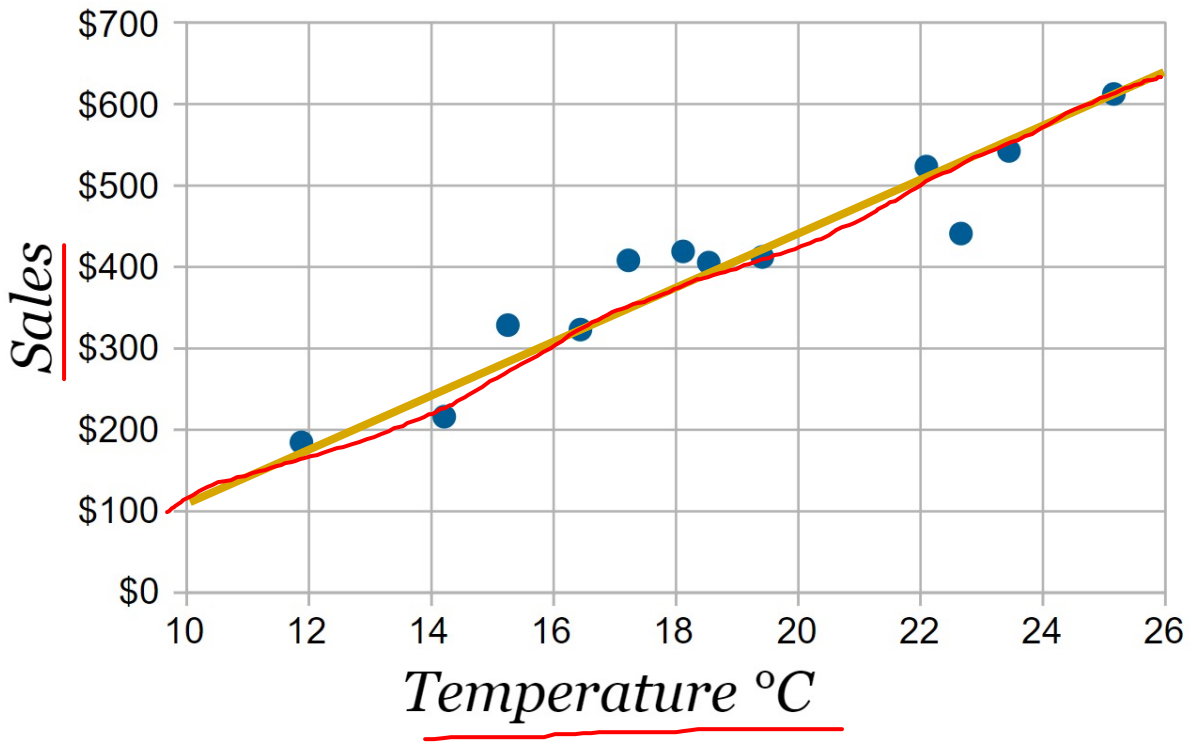
E.g. Which attribute is split on.

3) Choose the objective function:

E.g. Training error.

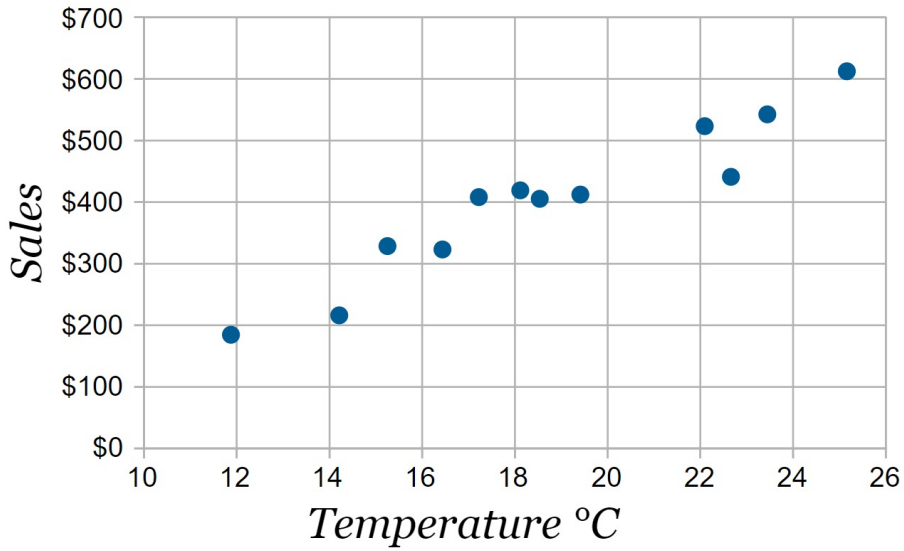
4) Learning:
Train the model to optimize the objective function.

Example: Regression



A learning system: basic cycle

1) **Data**: $D = \{d_1, d_2, ..., d_n\}$



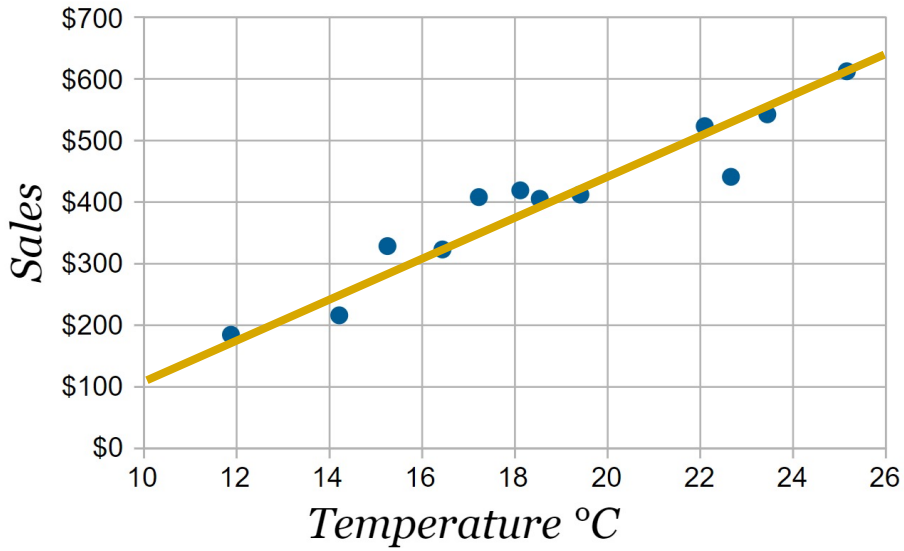
A learning system: basic cycle

2) Choose a learner:

Select a model with parameters (e.g. Regression)

Parameter: A model value learned in training

E.g. $f(x) = \boldsymbol{a}x + \boldsymbol{b}$ ← a and b



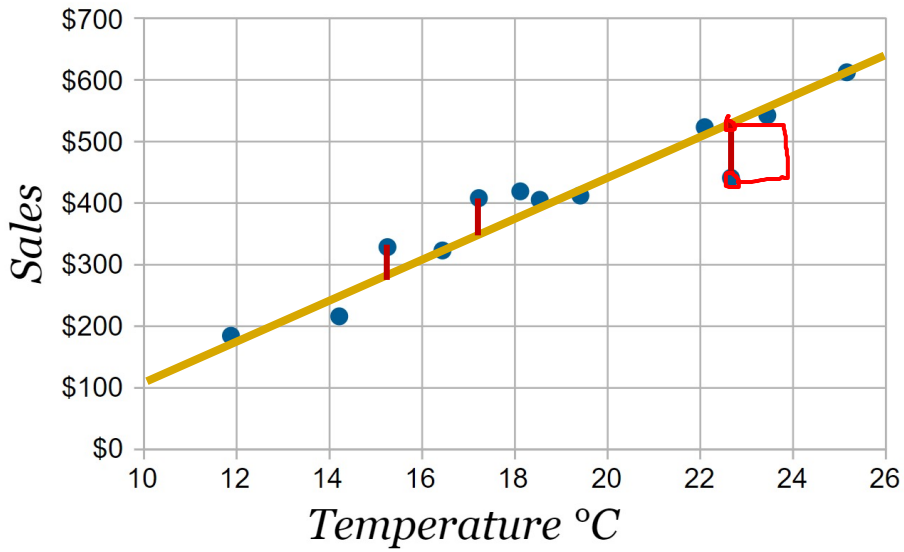
AI Academy

NC STATE

A learning system: basic cycle

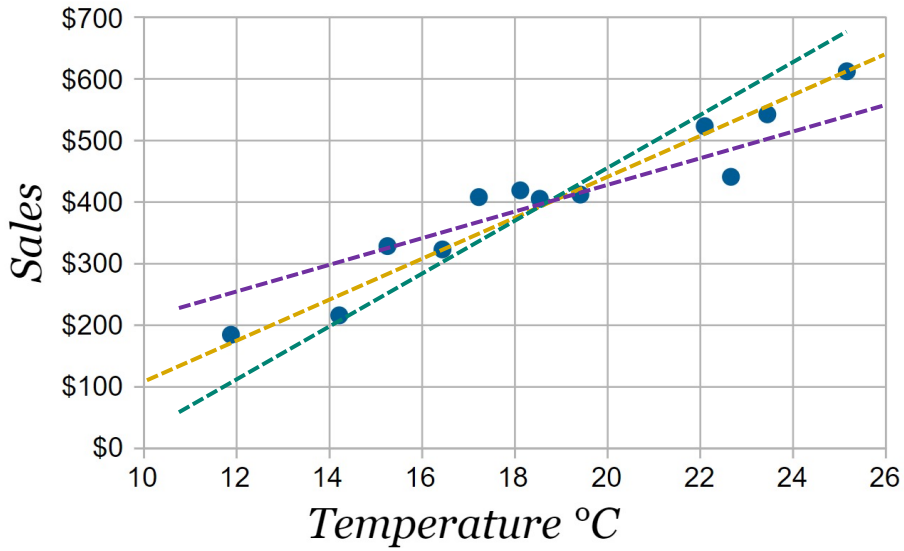
3) Choose the objective function
Squared Error:

$$\frac{1}{n}\sum\left(y - f\left(x\right)\right)^2$$



A learning system: basic cycle

4) Learning:
Train the model to optimize the objective function.



AI Academy

NC STATE

A learning system: basic cycle

1) Data: $D = \{d_1, d_2, .., d_n\}$

2) Choose a learner:

Select a model with parameters (e.g. Regression)

Parameter: A model value learned in training

E.g. $f(x) = \boldsymbol{a}x + \boldsymbol{b}$ ← a and b

3) Choose the objective function

Squared Error:

$$\frac{1}{n}\sum (y - f(x))^2$$

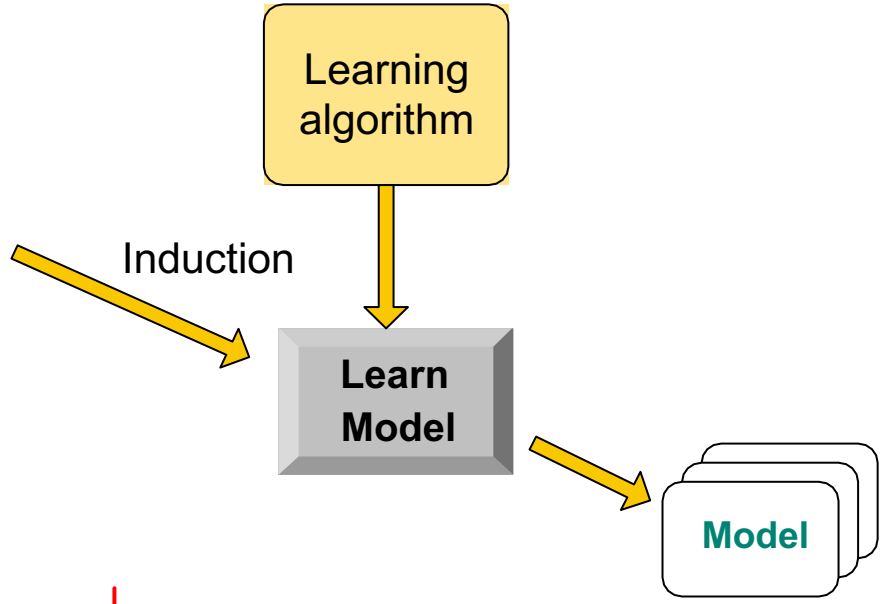
4) Learning:

Train the model to optimize the objective function

Illustrating Classification Task

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

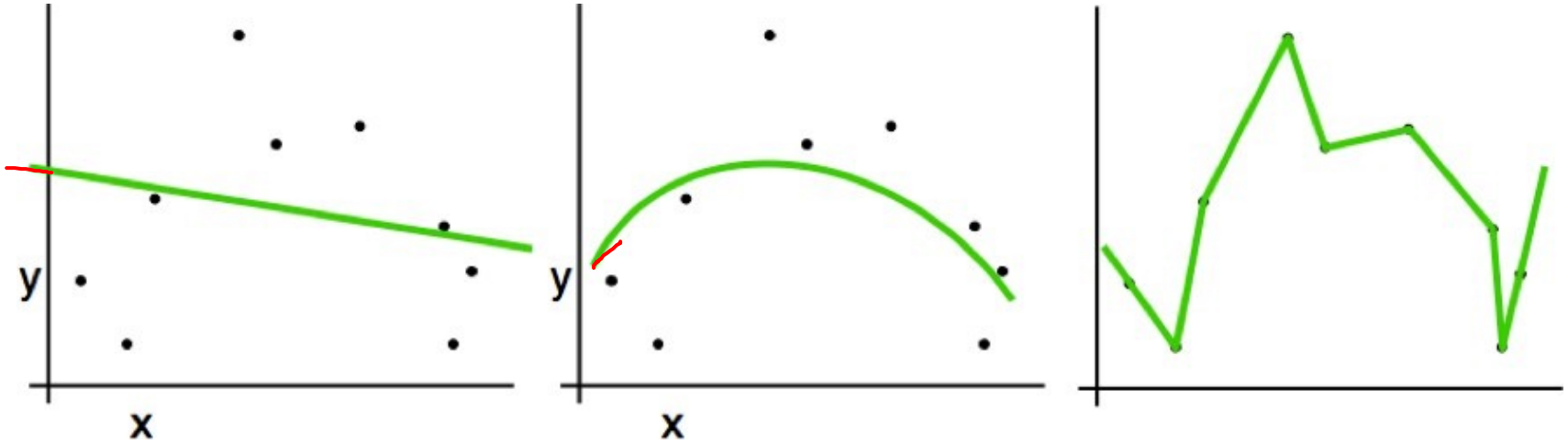
Training Set



Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

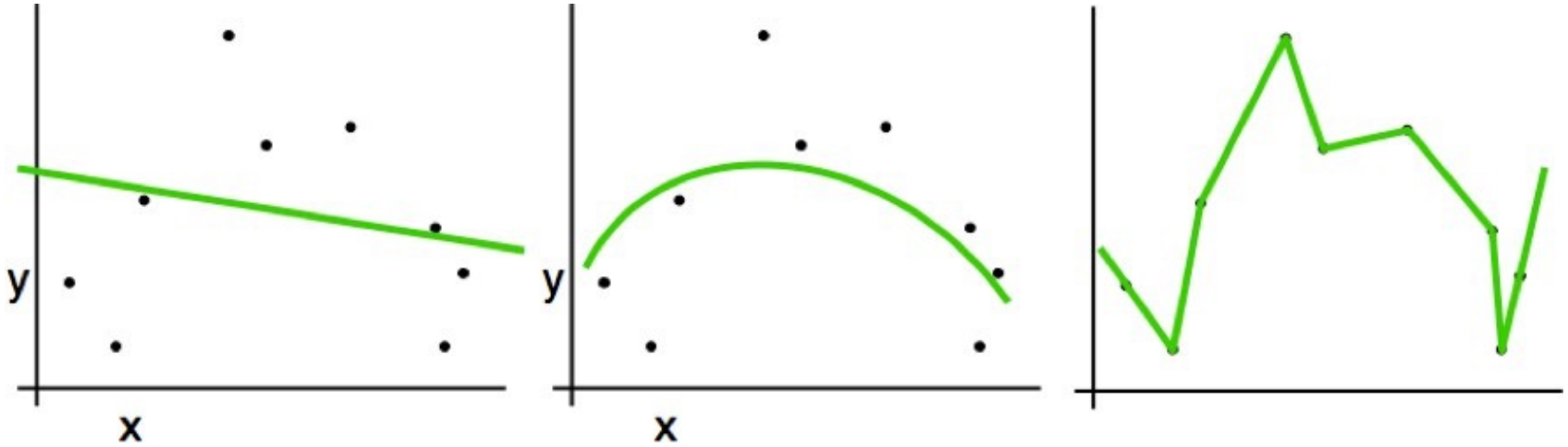
Test Set

Evaluate 3 Learners: Which is best?



(1) (2) (3)

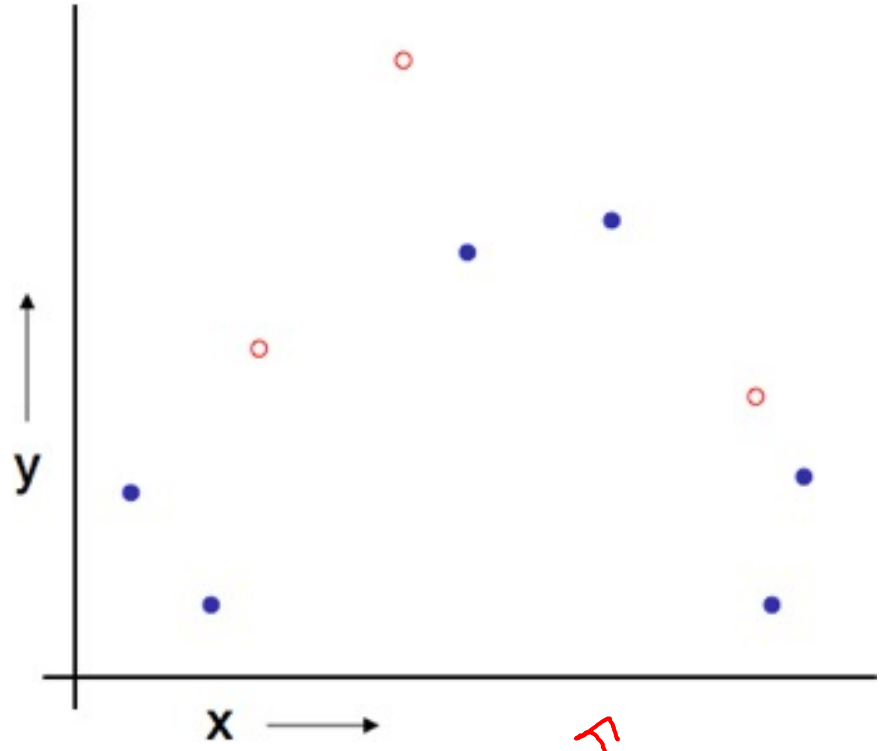
Evaluate 3 Learners: Which is best?



How well are you going to predict ***future*** data drawn from the same distribution?

Holdout Method

1. Randomly choose 30% of the data to be in the **test set.**
2. The remaining is in a **training set.**



Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Training Set

Induction

Learn Model

Model

Apply Model

Deduction

Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

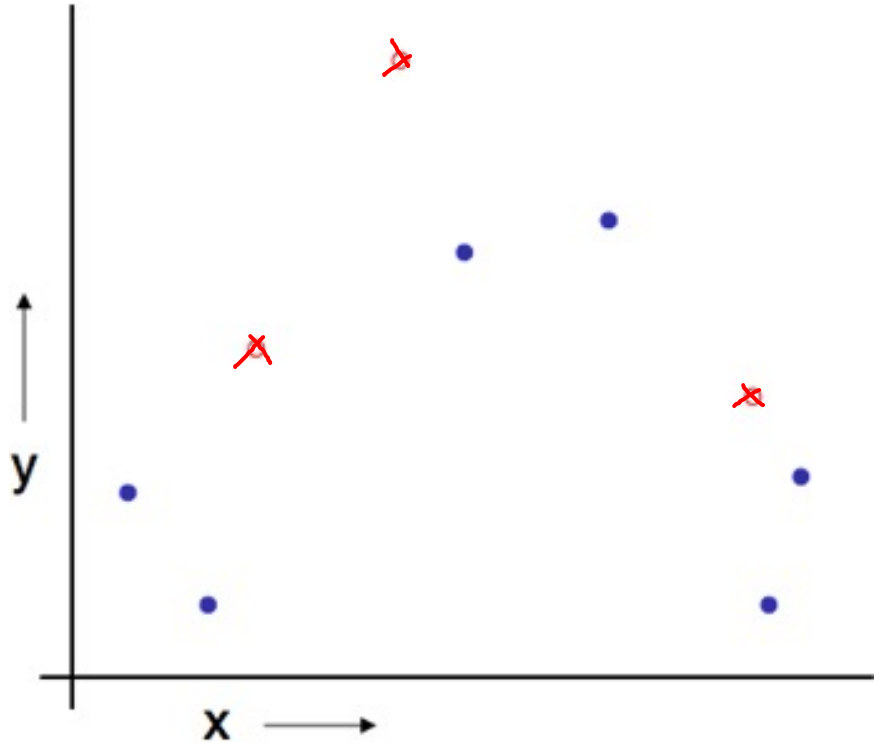
Test Set

AI Academy

NC STATE

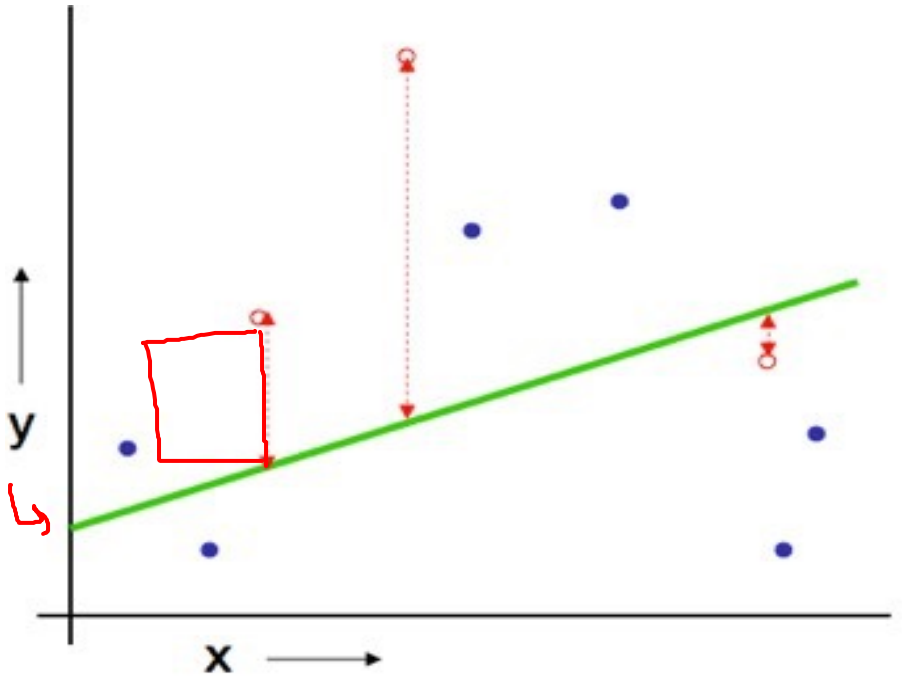
Holdout Method

1. Randomly choose 30% of the data to be in the **test set**
2. The remaining is in a **training set.**
3. Perform your regression on the training set.



Holdout Method

1. Randomly choose 30% of the data to be in the **test set.**
2. The remaining is in a **training set.**
3. Perform your regression on the training set.
4. Estimate your future performance with the test set.

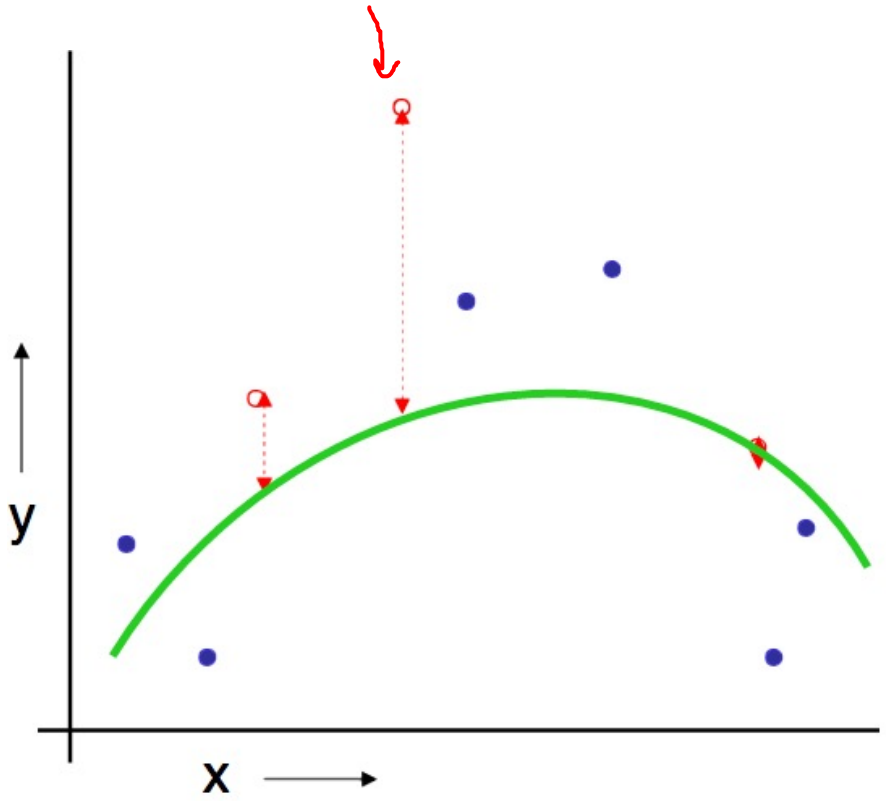


(Linear regression example)

Mean Squared Error = 2.4

Holdout Method

1. Randomly choose 30% of the data to be in the **test set.**
2. The remaining is in a **training set.**
3. Perform your regression on the training set.
4. Estimate your future performance with the test set.

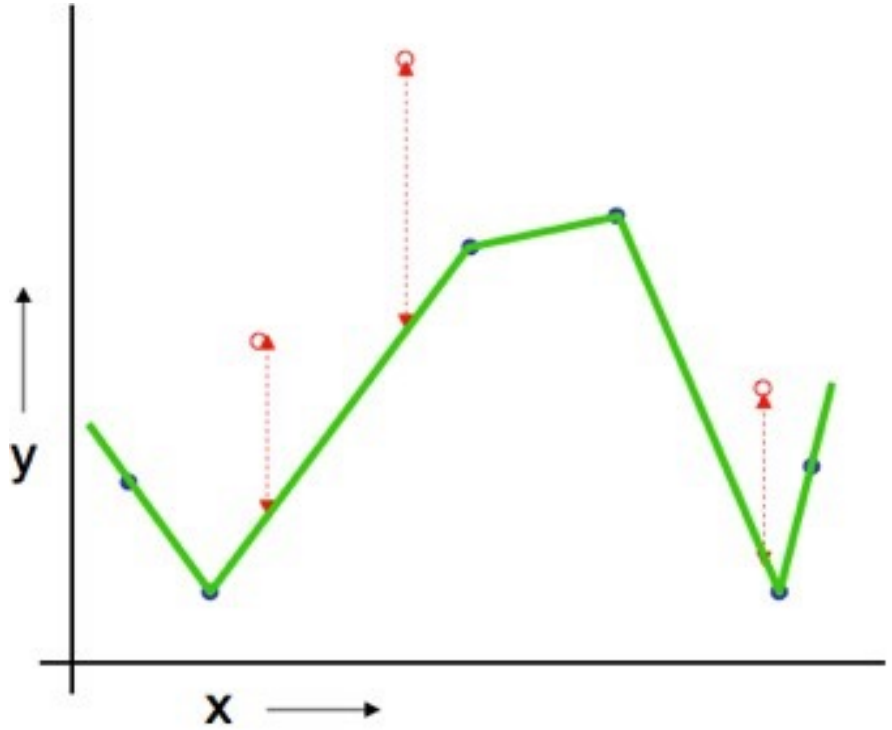


(Quadratic regression example)

Mean Squared Error = 0.9

Holdout Method

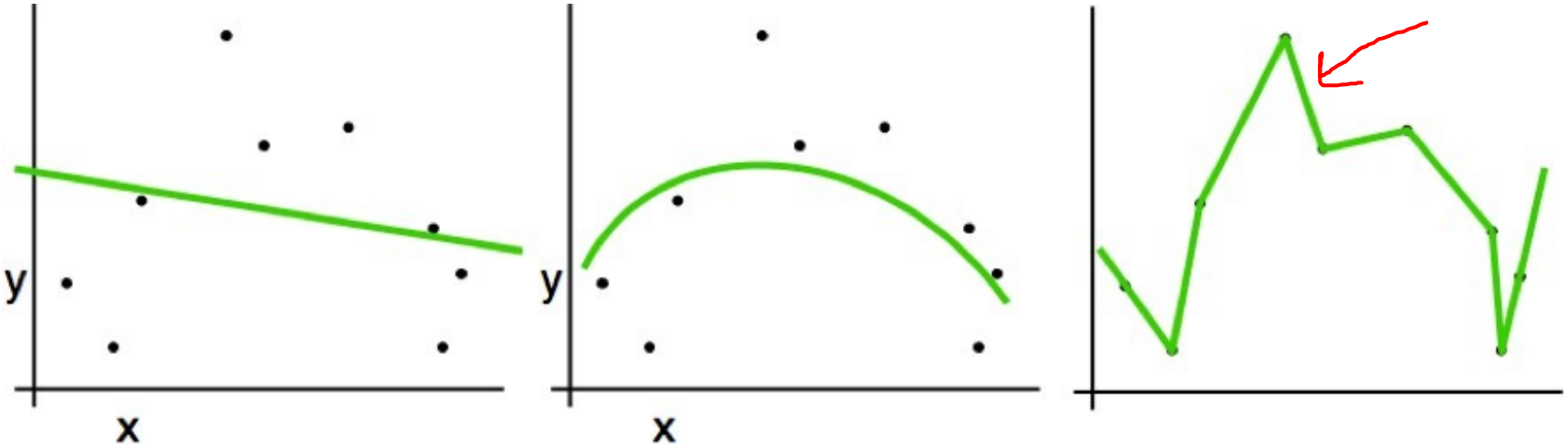
1. Randomly choose 30% of the data to be in the **test set.**
2. The remaining is in a **training set.**
3. Perform your regression on the training set.
4. Estimate your future performance with the test set.



(Join the dots example)

Mean Squared Error = 2.2

Evaluate 3 Learners: Which is best?



(1)
MSE = 2.4

(2)
MSE = 0.9

(3)
MSE = 2.2

MSE on **Test Data!**

Holdout Method

- **Good news**
  - Very very simple.
  - We choose the method best test-set score.

Holdout Method

- **Good news**
  - Very very simple.
  - We choose the method best test-set score.
- **Bad news**
  - **Wastes data**: Only 70% of data is used to train.
  - **Not Robust**: What if one model is just lucky on that test data?

Variations on Holdout

- **Repeated Resampling:**
  - Repeat holdout with random train/test splits.
- **Bootstrap:**
  - Repeated resampling *with replacement*.
- **Stratified Sampling:**
  - Oversampling vs Under sampling.

Repeated Resampling

1. Randomly sample a training/test split.
2. Evaluate the classifier.
3. Repeat *k* times and average the results.

- **Advantages**
  - Better estimate than a single split.
  - Can estimate the variance of error.
- **Disadvantages**
  - Results still depend on a random sample.
  - May take many rounds.

Stratification

- **Problem:** What if one class is rare?
  - Little training data → high error.
- **Solution:** Use stratified sampling.

- **Example:** dataset with two classes A and B.
  - **Goal:** Construct a 30% test set.
  - **Approach:** Sample 30% of:
    - Class A instances
    - Class B instance
- ***This procedure doesn't work well on small datasets.***

Learning Objectives: Validation Methods

You now should be able to:

- Evaluate a model using the hold-out approach and its variants.

AI Academy
NC STATE

Evaluation Measures

Exercises

AI Academy

Practice

- Use the **Holdout Method** to evaluate a “decision stump” (1-level decision tree). What is the **testing error**?

- Use the first 4 instances for training and the last 2 for test.

Color	Class
Red	**Yes**
Red	*No*
Red	**Yes**
Blue	**Yes**
Blue	**Yes**
Red	*No*