Instance Based Learning

05 August 2024 19:00

no explicit generaliz^n → instead → compare new instance w/ instances in training. → Training sample.
stored in memory.

no model created — machine stores data
- uses
  - pattern recog
  - clustering
  - anomaly detection
- then generalizes to new instances based on some similarity measure.

called "instance based" cuz hypothesis built from instances.

→ memory based learning / lazy learning
(processing delayed till new instance classified)
↓ time complexity?
O(n): depends on size of training data.

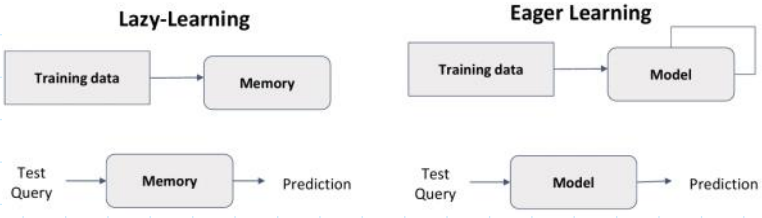
Adv
1. No need for model creation
2. Can handle small datasets
3. More flexibility

Disadv
1. Slow pred
2. Less accurate pred. → depends on training instances.
3. Limited understanding of data → no rel^n b/w I/P & O/P

Usual/Conventional Machine Learning	Instance Based Learning
Prepare the data for model training	Prepare the data for model training. No difference here
Train model from training data to estimate model parameters i.e. discover patterns	Do not train model. Pattern discovery postponed until scoring query received
Store the model in suitable form	There is no model to store
Generalize the rules in form of model, even before scoring instance is seen	No generalization before scoring. Only generalize for each scoring instance individually as and when seen
Predict for unseen scoring instance using model	Predict for unseen scoring instance using training data directly
Can throw away input/training data after model training	Input/training data must be kept since each query uses part or full set of training observations
Requires a known model form	May not have explicit model form
Storing models generally requires less storage	Storing training data generally requires more storage

Lazy v. Eager learning.

Lazy	Eager
1. Training data → Memory	1. Training data → Model
2. Test query to memory	2. Test query to model.
3. No global model created	3. Global model created.
4. For every test, common procedure on training data to find pred^n	4. Gen. model on training set, test query on global model.



Instance based ML algos:

1. **KNN**
   new instance classified based on majority class among k closest neighbors.
2. **SOM**
   unsupervised: uses NN to map high dimensional data to low dim^{n}al space.
   preserves topological properties of I/P.
3. **Learning Vector Quantization**
   supervised: creates prototypes that represent class regions.
   then adjust prototypes to better partition the space.
4. **Locally weighted learning.**
   weights training instances acc. to distance from query point.
   higher weight to closer instances.

higher weight to closer instances.

⑤ Case Based reasoning.
solves new problems by adapting solutions of old ones.