Cat boost.

↓

ML library which handles Categorical data automatically.

Out-of the box support for the more descriptive data formats.

Catboost implements Symetric trees.
(helps decreasing prediction time).

Default max-depth = 6

Dataset ordered in time
(Catboost creates an artificial time for each datapoint)

time	datapoint	label
	x_1	10
	x_2	12
	x_3	9
	x_4	4
	x_5	52
	x_6	22
	x_7	33
	x_8	34
	x_9	32
	x_{10}	12

10 points → 9 models.
✗ Computationaly expensive

log(no.of datapoints) models
→ log(10) = 2.7

Model trained on n data points is used to calculate the residuals

Ordered boosting.

Catboost divides a data into Random permutations (default = 4.) and applies ordered boosting.

bagging_temperature. → this parameter for randomness.

feature	label
x_1	10
x_3	15
x_1	20
x_3	15
x_1	30
x_2	20

response Coding.

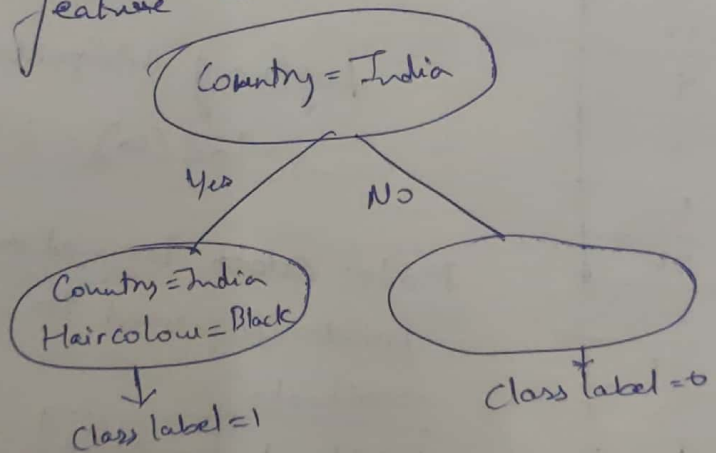
feature	label
20	10
15	15
20	20
15	15
20	30
20	20

This leads to target leakage

To avoid this, only datapoints that are past in time to a data point are considered.

Sometimes 0, → Use Laplace smoothing.

Categorical feature Combinations.



One-hot Encoding → one_hot_max_size = N.

Handling Numerical features → Same as other Algos.
(best split based on Information Gain).

Limitations

1) Not suitable for Sparse matrices.
2) Not suitable if it has many numerical features (Catboost takes more time to train than Light GBM).

T-SNE

T-distributed Stochastic Neighborhood Embedding.

Best for visualization of Data.

PCA: → Basic
MNIST
} → 2-dim.
Christopher Olah blog.

✗ MDS, Sammon mapping, Graph-based techniques ← 20 yrs.

✓ t-SNE → 2008, Geoffry Hinton.

d-dim → 2d & 3d
t-SNE
Clean visualization.

PCA: preserve global shape of data

t-SNE - local shape/struc. preserve.
(+ global struc. by changing parameter)

Neighborhood; Embedding D → [2] → same value

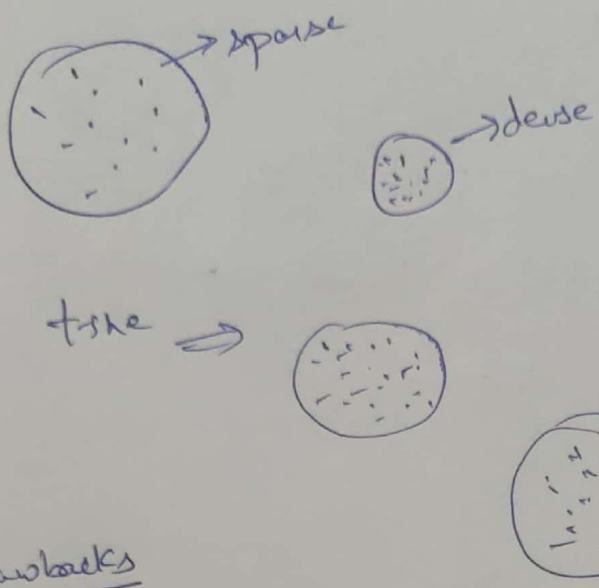
T-SNE is not a Deterministic algorithm

T-SNE is a Probabilistic algorithm

Every time you run T-SNE with a step, iteration, it might turn out different.

T-SNE
- → Nearest embedding
- → Stochastic
- → T-distributed to prevent Crowding.

t-SNE → expands dense clusters
Shrinks sparse clusters



Drawbacks

① One of the things you cannot read from t-sne is whether a cluster is dense or sparse.

i.e You cannot come to same density Conclusion using t-sne.

② t-sne does not preserve distance between clusters.

Thumb rule

① Run steps/iter till shapes stabilize

② perplexity $2 \leq p < n$

Never Run t-sne just Once and start reading it.

③ re-run t-sne with (p, step) multiple times.

Trajial knot

Epsilon → How fast you should chage from one iteration to another.