A CLOUD GURU

Feature Selection

Brock Tubre
INSTRUCTOR

Feature Selection

Selecting the most relevant features from your data to prevent over-complicating the analysis, resolving potential inaccuracies, and removes irrelevant features or repeated information.

Deciding what to keep and what to get rid of.

Feature Selection Example

Problem

Can old dogs learn new tricks?

ID	age	fur_color	born_month	born_month_num	breed	trick_learned
1	2	brown	July	7	terrier	0
2	4	white	August	8	shepard	1
3	3	golden	June	6	terrier	1
4	7	brown	September	8	collie	1
5	2	black	June	6	retriever	0
6	6	black	May	5	collie	1

Feature Selection Example

Problem

Can old dogs learn new tricks?

ID	age	fur_color	born_month	born_month_num	breed	trick_learned
1	2	brown	July	7	terrier	0
2	4	white	August	8	shepard	1
3	3	golden	June	6	terrier	1
4	7	brown	September	8	collie	1
5	2	black	June	6	retriever	0
6	6	black	May	5	collie	1

Feature Selection Example

Problem

Can old dogs learn new tricks?

ID	age	fur_color	born_month	born_month_num	breed	trick_learned
1	2	brown	July	7	terrier	0
2	4	white	August	8	shepard	1
3	3	golden	June	6	terrier	1
4	7	brown	September	8	collie	1
5	2	black	June	6	retriever	0
6	6	black	May	5	collie	1

Feature Selection Example

Problem

Can old dogs learn new tricks?

ID	age	fur_color	born_month	born_month_num	breed	trick_learned
1	2	brown	July	7	terrier	0
2	4	white	August	8	shepard	1
3	3	golden	June	6	terrier	1
4	7	brown	September	8	collie	1
5	2	black	June	6	retriever	0
6	6	black	May	5	collie	1

Feature Selection Example

Problem

Can old dogs learn new tricks?

ID	age	fur_color	born_month	born_month_num	breed	trick_learned
1	2	brown	July	7	terrier	0
2	4	white	August	8	shepard	1
3	3	golden	June	6	terrier	1
4	7	brown	September	8	collie	1
5	2	black	June	6	retriever	0
6	6	black	May	5	collie	1

Feature Selection Example

Problem

Can old dogs learn new tricks?

ID	age	fur_color	born_month	born_month_num	breed	trick_learned
1	2	brown	July	7	terrier	0
2	4	white	August	8	shepard	1
3	3	golden	June	6	terrier	1
4	7	brown	September	8	collie	1
5	2	black	June	6	retriever	0
6	6	black	May	5	collie	1

Feature Selection

Feature selection is an intuitive step humans take to reduce the number of features.

Principle Component Analysis (PCA)

Principle Component Analysis (PCA)

An unsupervised learning algorithm that reduces the number of features while still retaining as much information as possible.

Reduces the total number of features in a dataset.

Feature Selection Use Case

Problem	Technique	Why
Data is too large due to the large number of features	Principle Component Analysis (PCA)	Algorithm that reduces total number of features
Useless features that do not help solve ML problem	Feature Selection	Remove features that do not help solve the problem