DEPARTAMENTO DE SEÑALES, SISTEMAS Y RADIOCOMUNICACIONES

Machine Learning Lab

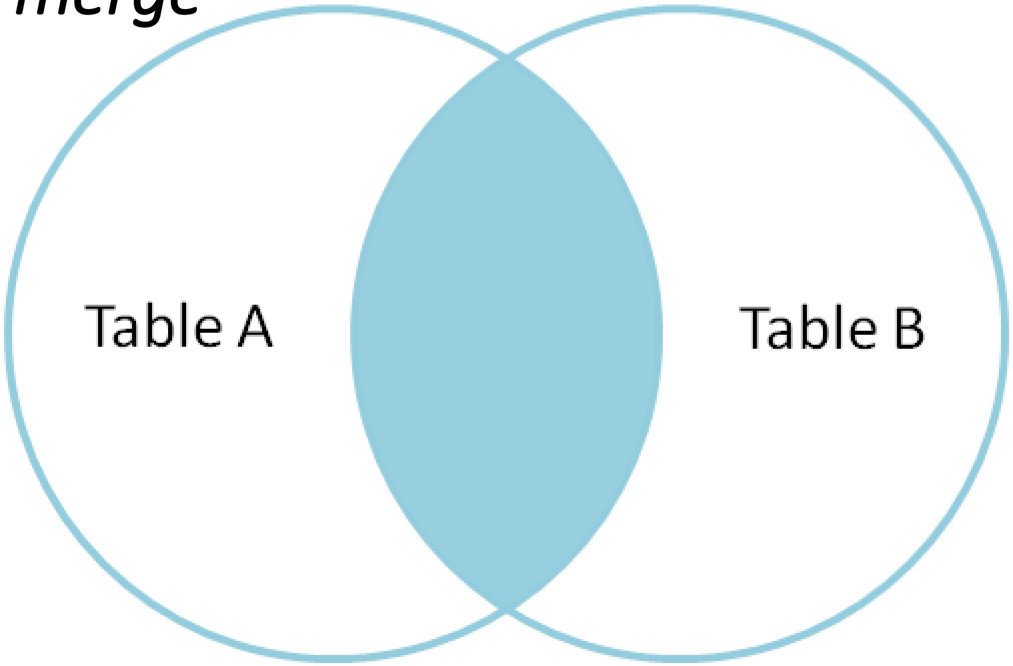
Master of Science in Signal Theory and Communications

TRACK: Signal Processing and Machine Learning for Big Data

Departamento de Señales, Sistemas y Radiocomunicaciones
E.T.S. Ingenieros de Telecomunicación
Universidad Politécnica de Madrid

Python Introduction: type of joins

Inner join : *merge*



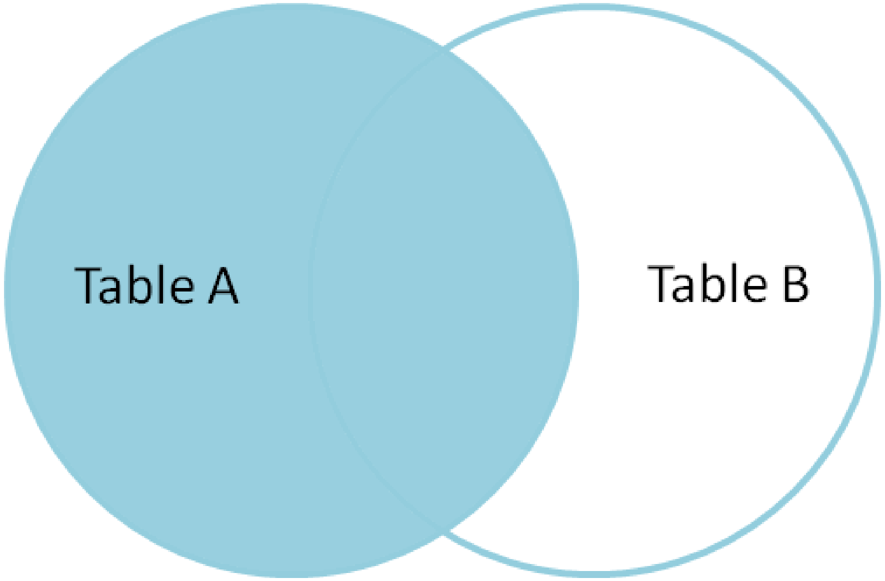
returns a new DataFrame that contains **only** those rows that have matching values in *both* of the original DataFrames.

Python Introduction: type of joins

```
Merge_df = pd.merge(left_df, right_df,
on=‘column', how='inner')
```

Python Introduction: type of joins

Left join : *merge + how=‘left’*



- Return all of the rows from the left DataFrame, even those rows whose join key(s) do not have values in the right DataFrame.

- Rows in the left DataFrame that are missing values for the join key(s) in the right DataFrame will simply have null (i.e., NaN or None)

Note: a left join will still discard rows from the right DataFrame that do not have values for the join key(s) in the left DataFrame.

Python Introduction

Other join types

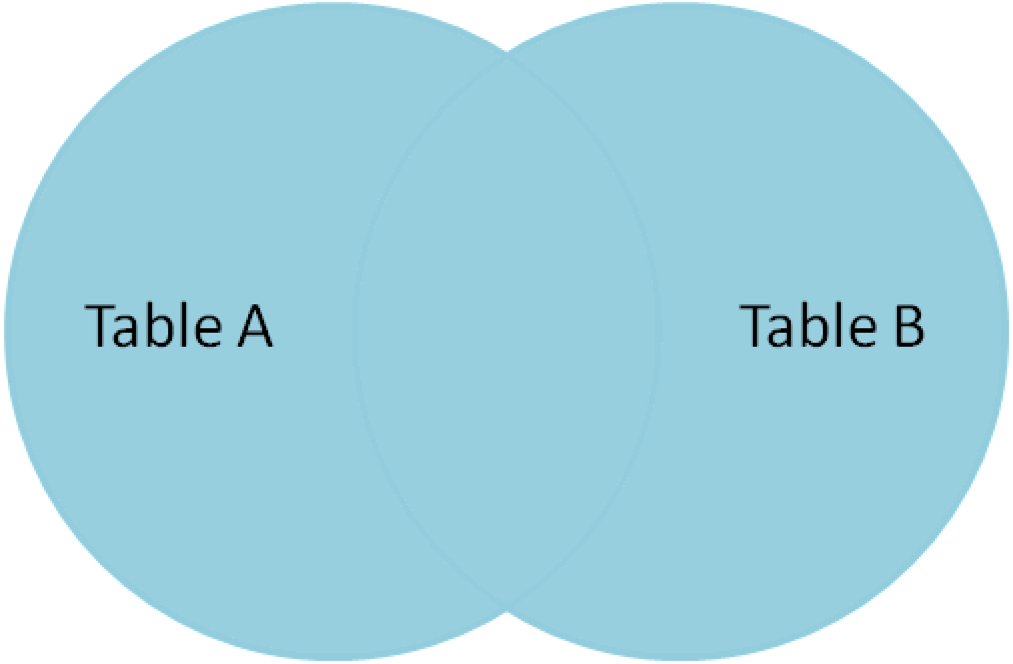
The pandas merge function supports other join types:

- **Right (outer) join**: Invoked by passing how='right' as an argument. Similar to a left join, except all rows from the right DataFrame are kept, while rows from the left DataFrame without matching join key(s) values are discarded.

.

Python Introduction

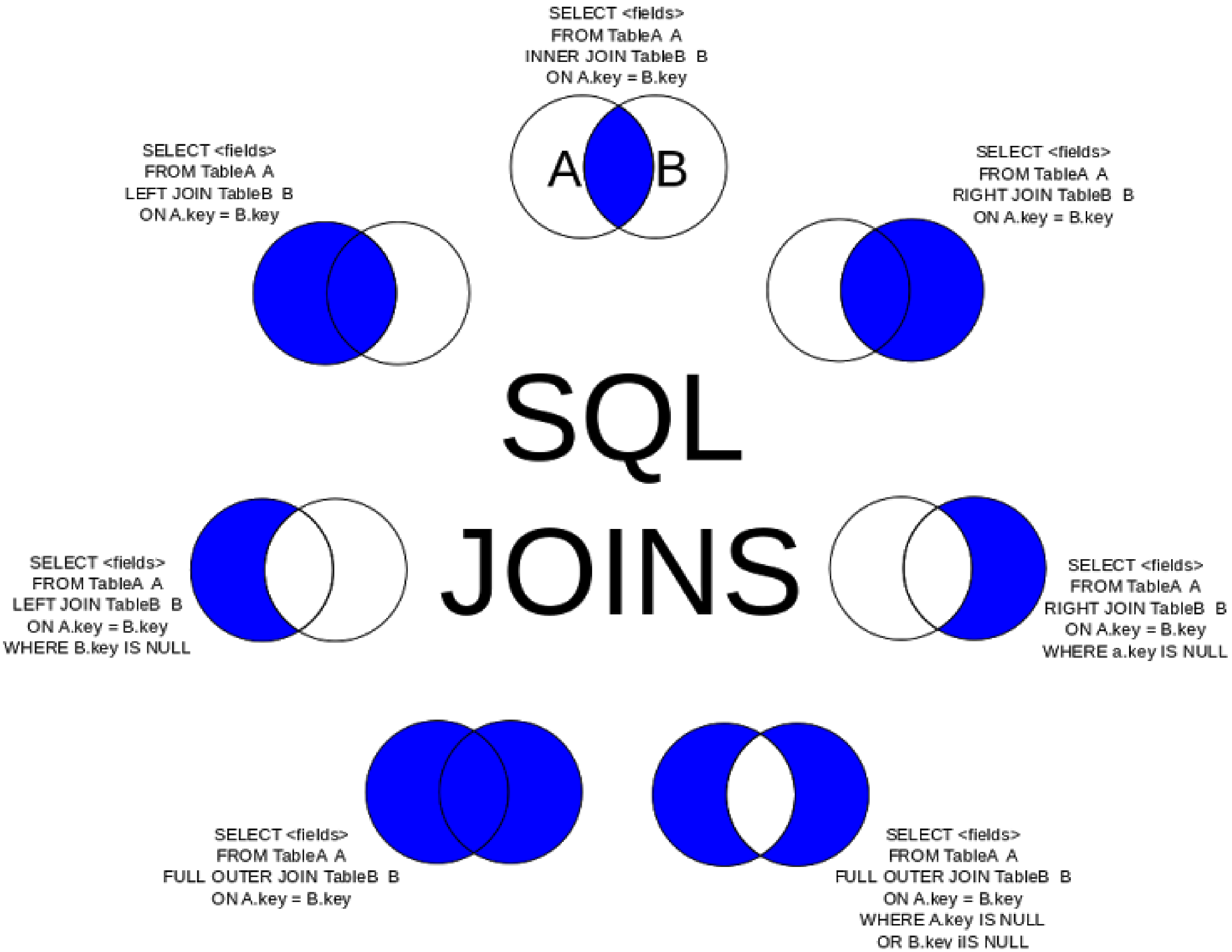
- **Full (outer) join**: Invoked by passing how='outer' as an argument. This join type returns the all pairwise combinations of rows from both DataFrames; i.e., the result DataFrame will NaN where data is missing in one of the dataframes. This join type is very rarely used.



Python Introduction: join and merge Pandas dataframes

See, for example:

https://chrisalbon.com/python/data_wrangling/pandas_join_merge_dataframe/



SQL JOINS

```
SELECT <fields>
FROM TableA A
INNER JOIN TableB B
ON A.key = B.key
```

```
SELECT <fields>
FROM TableA A
LEFT JOIN TableB B
ON A.key = B.key
```

```
SELECT <fields>
FROM TableA A
RIGHT JOIN TableB B
ON A.key = B.key
```

```
SELECT <fields>
FROM TableA A
LEFT JOIN TableB B
ON A.key = B.key
WHERE B.key IS NULL
```

```
SELECT <fields>
FROM TableA A
RIGHT JOIN TableB B
ON A.key = B.key
WHERE a.key IS NULL
```

```
SELECT <fields>
FROM TableA A
FULL OUTER JOIN TableB B
ON A.key = B.key
```

```
SELECT <fields>
FROM TableA A
FULL OUTER JOIN TableB B
ON A.key = B.key
WHERE A.key IS NULL
OR B.key iIS NULL
```

SQL
VS
Pandas

Since many potential pandas users have some familiarity with SQL, this page is meant to provide some examples of how various SQL operations would be performed using pandas.

https://pandas.pydata.org/pandas-docs/stable/getting_started/comparison/comparison_with_sql.html