Breakdown of sklearn.datasets.make_regression

Works to create synthetic datasets to test regression algorithms, for educational purposes, benchmarking machine learning models, or simulating data scenarios.

1. **n_samples** (int, default=100):
   - Sample Size
   - Description: The number of samples to generate.
   - Example: If I want to simulate a study with 1000 individuals, set `n_samples=1000`.

2. **n_features** (int, default=100):
   - Number of Predictor Variables
   - Description: The number of features (independent variables) in the dataset.
   - Example: If I was simulating a situation with 5 different predictors (like age, income, etc.), I would set `n_features=5`.

3. **n_informative** (int, default=10):
   - Number of Informative Features
   - Description: The number of features actually used to build the linear model.
   - Example: If only 3 out of 5 features are relevant to my prediction, I would set `n_informative=3`.

4. **n_targets** (int, default=1):
   - Number of Response Variables
   - Description: The number of targets. If this is greater than 1, then it's a multi-output regression.
   - Example: For predicting a single outcome like house price, I would use `n_targets=1`. For predicting both house price and time on the market, I would use `n_targets=2`.

5. **bias** (float, default=0.0):
   - Intercept
   - Description: The bias term in the linear model.
   - Example: If I know my model should have an intercept of 10, I would set `bias=10`.

6. **effective_rank** (int or None, default=None):
   - Rank of the coefficient matrix
   - Description: If not None, the number of singular vectors required to explain the data. This is the correlation among the features.
   - Example: For a low-rank scenario with super correlated features, I would set `effective_rank` to a value lower than `n_features`

7. **tail_strength** (float, default=0.5):
   - Tail Strength of Singular Values

8. **noise** (float, default=0.0):
   - Measurement Error / Noise
   - The standard deviation of the Gaussian (normal) noise applied to the output.
   - If my outcome measurements have a noise level with a standard deviation of 3, I would set `noise=3`.

9. **shuffle** (bool, default=True):
   - Randomization
   - Description: Shuffle the samples and the features.

10. **coef** (bool, default=False):
    - Returning the Coefficient
    - Description: If True, the coefficients of the underlying linear model are returned.
    - Example: To inspect the coefficients that were used to generate the data, set `coef=True`.

11. **random_state** (int, RandomState instance or None, default=None):
    - The random seed
    - Example: For reproducible results, set `random_state` to a fixed number like `random_state=123`.