regression_forest_testing

December 14, 2025

1 Testing Notebook - Regression Random Forest

[1]:

```
from scripts.regression_random_forest import RandomForestRegressor
from sklearn.ensemble import RandomForestRegressor as␣
  ↪SklearnRandomForestRegressor
from sklearn.model_selection import train_test_split, cross_validate
from sklearn.datasets import make_regression
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd


RANDOM_STATE = 42
```

1.1 Helper functions

[2]:

```
"""Generate synthetic regression data using scikit-learn's make_regression␣
  ↪function."""
def generate_regression_data(n_samples=100, n_features=1, noise=0.1,␣
  ↪random_state=RANDOM_STATE):
  X, y = make_regression(n_samples=n_samples, n_features=n_features,␣
  ↪noise=noise, random_state=random_state)
  print("Generated regression data with {} samples and {} features.".
  ↪format(n_samples, n_features))
  print("First 5 samples of X:\n", X[:5])
  print("First 5 samples of y:\n", y[:5])
  return X, y

"""Evaluate a regression model using cross-validation and return performance␣
  ↪metrics."""
def evaluate_model(model, X, y, cv=5):
    cv_results = cross_validate(
        model,
        X,
        y,
        cv=cv,
        scoring=["r2", "neg_mean_squared_error", "neg_mean_absolute_error"],
        return_train_score=False,
    )

    return {
        "r2_mean": np.mean(cv_results["test_r2"]),
        "r2_std":  np.std(cv_results["test_r2"]),
        "mse_mean": -np.mean(cv_results["test_neg_mean_squared_error"]),
        "mae_mean": -np.mean(cv_results["test_neg_mean_absolute_error"]),
    }
```

1.2 Definition of testing configurations

[3]:

```
TESTING_CONFIGS = [
  {
    "description": "Basic regression data with low noise",
    "n_samples": 100,
    "n_features": 1,
    "noise": 0.1
  },
  {
    "description": "Regression data with higher noise",
    "n_samples": 100,
    "n_features": 1,
    "noise": 10.0
  },
  {
    "description": "Regression data with multiple features",
    "n_samples": 200,
    "n_features": 5,
    "noise": 5.0
  },
  {
    "description": "Larger dataset with moderate noise",
    "n_samples": 5000,
    "n_features": 3,
    "noise": 2.0
  },
  {
    "description": "Large dataset with low noise",
    "n_samples": 10000,
    "n_features": 2,
    "noise": 0.5
  },
  {
    "description": "Large dataset with high noise",
    "n_samples": 10000,
    "n_features": 2,
    "noise": 20.0
  },
  {
    "description": "Moderate dataset with many features",
    "n_samples": 1000,
    "n_features": 50,
    "noise": 5.0
  }
]

HYPERPARAM_CONFIGS = [
    {
        "name": "rf_shallow",
        "n_estimators": 100,
        "max_depth": 2,
        "min_samples_leaf": 1,
        "min_samples_split": 2,
        "max_features": "sqrt",
        "bootstrap": True,
    },
    {
        "name": "rf_medium",
        "n_estimators": 100,
        "max_depth": 4,
        "min_samples_leaf": 1,
        "min_samples_split": 2,
        "max_features": "sqrt",
        "bootstrap": True,
    },
    {
        "name": "rf_deep",
        "n_estimators": 100,
        "max_depth": 8,
        "min_samples_leaf": 1,
        "min_samples_split": 2,
        "max_features": "sqrt",
        "bootstrap": True,
    },
    {
        "name": "rf_unrestricted",
        "n_estimators": 100,
        "max_depth": None,
        "min_samples_leaf": 1,
        "min_samples_split": 2,
        "max_features": "sqrt",
        "bootstrap": True,
    },
    {
        "name": "rf_regularized_leaf",
        "n_estimators": 100,
        "max_depth": None,
        "min_samples_leaf": 5,
        "min_samples_split": 2,
        "max_features": "sqrt",
        "bootstrap": True,
    },
    {
        "name": "rf_strongly_regularized",
        "n_estimators": 100,
        "max_depth": None,
        "min_samples_leaf": 10,
        "min_samples_split": 10,
        "max_features": "sqrt",
        "bootstrap": True,
    },
    {
        "name": "rf_bagging_only",
        "n_estimators": 100,
        "max_depth": None,
        "min_samples_leaf": 1,
        "min_samples_split": 2,
        "max_features": None,
        "bootstrap": True,
  }
]
```

1.3 Runing the tests

For each configuration we will train the scikit-learn's built-in regression random forest model as well as our implementation, we collect useful statistics and compare them.

[4]:

```
results = []

for config in TESTING_CONFIGS:
    print("=" * 80)
    print(f"Dataset: {config['description']}")
    print("=" * 80)

    X, y = generate_regression_data(
        n_samples=config["n_samples"],
        n_features=config["n_features"],
        noise=config["noise"],
    )
```

```python
    for hp in HYPERPARAM_CONFIGS:
        print("-" * 80)
        print(f"Hyperparams: {hp['name']}")

        model_specs = [
            (
                "our",
                RandomForestRegressor(
                    n_estimators=hp["n_estimators"],
                    max_depth=hp["max_depth"],
                    min_samples_leaf=hp["min_samples_leaf"],
                    min_samples_split=hp["min_samples_split"],
                    max_features=hp["max_features"],
                    bootstrap=hp["bootstrap"],
                    random_state=RANDOM_STATE
                ),
            ),
            (
                "sklearn",
                SklearnRandomForestRegressor(
                    n_estimators=hp["n_estimators"],
                    max_depth=hp["max_depth"],
                    min_samples_leaf=hp["min_samples_leaf"],
                    min_samples_split=hp["min_samples_split"],
                    max_features=hp["max_features"],
                    bootstrap=hp["bootstrap"],
                    random_state=RANDOM_STATE
                ),
            ),
        ]

        for model_name, model in model_specs:
            metrics = evaluate_model(model, X, y)

            print(
                f"{model_name:8s} | "
                f"R²: {metrics['r2_mean']:.3f} ± {metrics['r2_std']:.3f} | "
                f"MSE: {metrics['mse_mean']:.3f} | "
                f"MAE: {metrics['mae_mean']:.3f}"
            )

            results.append({
                "dataset": config["description"],
                "hyperparams": hp["name"],
                "model": model_name,
                **metrics,
            })
```

```
================================================================================
Dataset: Basic regression data with low noise
================================================================================
Generated regression data with 100 samples and 1 features.
First 5 samples of X:
 [[ 0.93128012]
 [ 0.08704707]
 [-1.05771093]
 [ 0.31424733]
 [-0.47917424]]
First 5 samples of y:
 [ 38.9917296    3.4964533  -44.05770173  13.09112657 -19.9786311 ]
--------------------------------------------------------------------------------
Hyperparams: rf_shallow
our      | R²: 0.941 ± 0.029 | MSE: 83.496 | MAE: 6.466
sklearn  | R²: 0.941 ± 0.028 | MSE: 82.845 | MAE: 6.230
--------------------------------------------------------------------------------
Hyperparams: rf_medium
our      | R²: 0.990 ± 0.009 | MSE: 15.910 | MAE: 2.156
sklearn  | R²: 0.990 ± 0.009 | MSE: 15.183 | MAE: 2.099
--------------------------------------------------------------------------------
Hyperparams: rf_deep
our      | R²: 0.993 ± 0.009 | MSE: 11.851 | MAE: 1.334
sklearn  | R²: 0.993 ± 0.009 | MSE: 11.250 | MAE: 1.274
--------------------------------------------------------------------------------
Hyperparams: rf_unrestricted
our      | R²: 0.993 ± 0.009 | MSE: 11.850 | MAE: 1.334
sklearn  | R²: 0.993 ± 0.009 | MSE: 11.250 | MAE: 1.274
--------------------------------------------------------------------------------
Hyperparams: rf_regularized_leaf
our      | R²: 0.985 ± 0.018 | MSE: 25.583 | MAE: 2.354
sklearn  | R²: 0.970 ± 0.025 | MSE: 48.628 | MAE: 3.943
--------------------------------------------------------------------------------
Hyperparams: rf_strongly_regularized
our      | R²: 0.949 ± 0.029 | MSE: 75.613 | MAE: 5.365
sklearn  | R²: 0.915 ± 0.026 | MSE: 122.739 | MAE: 7.441
--------------------------------------------------------------------------------
Hyperparams: rf_bagging_only
our      | R²: 0.993 ± 0.009 | MSE: 11.850 | MAE: 1.334
sklearn  | R²: 0.993 ± 0.009 | MSE: 11.250 | MAE: 1.274
================================================================================
Dataset: Regression data with higher noise
================================================================================
Generated regression data with 100 samples and 1 features.
First 5 samples of X:
 [[ 0.93128012]
 [ 0.08704707]
 [-1.05771093]
 [ 0.31424733]
 [-0.47917424]]
First 5 samples of y:
 [ 50.77992943 -10.06527016 -34.91839191  10.52674299 -17.73837724]
--------------------------------------------------------------------------------
Hyperparams: rf_shallow
our      | R²: 0.877 ± 0.068 | MSE: 192.074 | MAE: 10.672
sklearn  | R²: 0.876 ± 0.070 | MSE: 192.633 | MAE: 10.757
--------------------------------------------------------------------------------
Hyperparams: rf_medium
our      | R²: 0.925 ± 0.030 | MSE: 116.684 | MAE: 8.703
sklearn  | R²: 0.926 ± 0.030 | MSE: 114.503 | MAE: 8.609
--------------------------------------------------------------------------------
Hyperparams: rf_deep
our      | R²: 0.914 ± 0.032 | MSE: 133.687 | MAE: 9.084
sklearn  | R²: 0.915 ± 0.032 | MSE: 132.492 | MAE: 9.097
--------------------------------------------------------------------------------
Hyperparams: rf_unrestricted
our      | R²: 0.914 ± 0.032 | MSE: 134.203 | MAE: 9.105
sklearn  | R²: 0.915 ± 0.032 | MSE: 132.799 | MAE: 9.111
--------------------------------------------------------------------------------
Hyperparams: rf_regularized_leaf
our      | R²: 0.919 ± 0.034 | MSE: 129.923 | MAE: 8.986
sklearn  | R²: 0.908 ± 0.039 | MSE: 150.154 | MAE: 9.487
--------------------------------------------------------------------------------
Hyperparams: rf_strongly_regularized
our      | R²: 0.888 ± 0.052 | MSE: 181.065 | MAE: 10.128
sklearn  | R²: 0.841 ± 0.076 | MSE: 258.953 | MAE: 12.212
--------------------------------------------------------------------------------
Hyperparams: rf_bagging_only
our      | R²: 0.914 ± 0.032 | MSE: 134.203 | MAE: 9.105
sklearn  | R²: 0.915 ± 0.032 | MSE: 132.799 | MAE: 9.111
================================================================================
Dataset: Regression data with multiple features
================================================================================
Generated regression data with 200 samples and 5 features.
First 5 samples of X:
 [[-0.3853136   0.1990597  -0.60021688  0.46210347  0.06980208]
 [ 0.13074058  1.6324113  -1.43014138 -1.24778318 -0.44004449]
 [-0.77300978  0.22409248  0.0125924  -0.40122047  0.0976761 ]
 [-0.57677133 -0.05023811 -0.23894805  0.27045683 -0.90756366]
 [-0.57581824  0.6141167   0.75750771 -0.2209696  -0.53050115]]
First 5 samples of y:
 [ -23.06825766 -124.34606251    6.99640504  -71.77655611   17.06336023]
--------------------------------------------------------------------------------
Hyperparams: rf_shallow
our      | R²: 0.612 ± 0.034 | MSE: 3556.656 | MAE: 44.420
sklearn  | R²: 0.606 ± 0.044 | MSE: 3609.998 | MAE: 44.611
```

```
--------------------------------------------------------------------------------
Hyperparams: rf_medium
our      | R²: 0.787 ± 0.046 | MSE: 1949.748 | MAE: 32.193
sklearn  | R²: 0.786 ± 0.029 | MSE: 1966.775 | MAE: 32.803
--------------------------------------------------------------------------------
Hyperparams: rf_deep
our      | R²: 0.876 ± 0.036 | MSE: 1135.633 | MAE: 23.644
sklearn  | R²: 0.876 ± 0.037 | MSE: 1143.984 | MAE: 24.194
--------------------------------------------------------------------------------
Hyperparams: rf_unrestricted
our      | R²: 0.880 ± 0.034 | MSE: 1103.044 | MAE: 23.330
sklearn  | R²: 0.876 ± 0.034 | MSE: 1145.143 | MAE: 23.624
--------------------------------------------------------------------------------
Hyperparams: rf_regularized_leaf
our      | R²: 0.841 ± 0.037 | MSE: 1473.872 | MAE: 26.523
sklearn  | R²: 0.813 ± 0.026 | MSE: 1710.348 | MAE: 28.939
--------------------------------------------------------------------------------
Hyperparams: rf_strongly_regularized
our      | R²: 0.793 ± 0.028 | MSE: 1909.516 | MAE: 30.493
sklearn  | R²: 0.746 ± 0.023 | MSE: 2340.845 | MAE: 33.863
--------------------------------------------------------------------------------
Hyperparams: rf_bagging_only
our      | R²: 0.914 ± 0.024 | MSE: 782.231 | MAE: 19.692
sklearn  | R²: 0.912 ± 0.029 | MSE: 801.749 | MAE: 20.101
================================================================================
Dataset: Larger dataset with moderate noise
================================================================================
Generated regression data with 5000 samples and 3 features.
First 5 samples of X:
 [[ 0.67796997 -1.28472777 -0.33102433]
 [ 1.03138053  0.3881858  -0.97027133]
 [-1.21689671  1.36337651 -0.60515624]
 [-0.54429615 -0.50442268 -1.5198928 ]
 [ 0.2074888   0.44567791  0.42350787]]
First 5 samples of y:
 [  -3.90976845  -46.39345674  -98.15873956 -157.1184525    47.56364341]
--------------------------------------------------------------------------------
Hyperparams: rf_shallow
our      | R²: 0.502 ± 0.014 | MSE: 4625.518 | MAE: 49.928
sklearn  | R²: 0.535 ± 0.011 | MSE: 4315.775 | MAE: 47.784
--------------------------------------------------------------------------------
Hyperparams: rf_medium
our      | R²: 0.781 ± 0.017 | MSE: 2038.090 | MAE: 32.157
sklearn  | R²: 0.794 ± 0.012 | MSE: 1915.485 | MAE: 31.155
--------------------------------------------------------------------------------
Hyperparams: rf_deep
our      | R²: 0.950 ± 0.005 | MSE: 465.040 | MAE: 14.148
sklearn  | R²: 0.953 ± 0.004 | MSE: 437.215 | MAE: 13.600
```

```
--------------------------------------------------------------------------------
Hyperparams: rf_unrestricted
our      | R²: 0.986 ± 0.003 | MSE: 133.930 | MAE: 6.248
sklearn  | R²: 0.986 ± 0.003 | MSE: 132.733 | MAE: 6.309
--------------------------------------------------------------------------------
Hyperparams: rf_regularized_leaf
our      | R²: 0.979 ± 0.005 | MSE: 196.901 | MAE: 7.593
sklearn  | R²: 0.974 ± 0.006 | MSE: 239.974 | MAE: 8.378
--------------------------------------------------------------------------------
Hyperparams: rf_strongly_regularized
our      | R²: 0.970 ± 0.006 | MSE: 280.350 | MAE: 9.148
sklearn  | R²: 0.962 ± 0.008 | MSE: 349.487 | MAE: 10.378
--------------------------------------------------------------------------------
Hyperparams: rf_bagging_only
our      | R²: 0.997 ± 0.000 | MSE: 23.490 | MAE: 2.804
sklearn  | R²: 0.998 ± 0.001 | MSE: 22.924 | MAE: 2.795
================================================================================
Dataset: Large dataset with low noise
================================================================================
Generated regression data with 10000 samples and 2 features.
First 5 samples of X:
 [[-0.5691482   1.59040357]
 [ 0.73487779  0.49097495]
 [ 0.20069869  1.10623156]
 [-1.07774393 -0.54427443]
 [ 0.10507597 -0.88502863]]
First 5 samples of y:
 [ -47.49496686   75.02051971   25.36349671 -108.87452141    5.49344142]
--------------------------------------------------------------------------------
Hyperparams: rf_shallow
our      | R²: 0.700 ± 0.008 | MSE: 2887.820 | MAE: 39.491
sklearn  | R²: 0.717 ± 0.006 | MSE: 2718.175 | MAE: 38.175
--------------------------------------------------------------------------------
Hyperparams: rf_medium
our      | R²: 0.911 ± 0.005 | MSE: 853.817 | MAE: 20.397
sklearn  | R²: 0.915 ± 0.005 | MSE: 820.737 | MAE: 20.132
--------------------------------------------------------------------------------
Hyperparams: rf_deep
our      | R²: 0.992 ± 0.002 | MSE: 78.464 | MAE: 5.565
sklearn  | R²: 0.991 ± 0.002 | MSE: 85.436 | MAE: 5.835
--------------------------------------------------------------------------------
Hyperparams: rf_unrestricted
our      | R²: 0.999 ± 0.001 | MSE: 13.858 | MAE: 1.367
sklearn  | R²: 0.999 ± 0.001 | MSE: 14.386 | MAE: 1.383
--------------------------------------------------------------------------------
Hyperparams: rf_regularized_leaf
our      | R²: 0.998 ± 0.001 | MSE: 22.540 | MAE: 1.731
sklearn  | R²: 0.997 ± 0.001 | MSE: 27.004 | MAE: 1.945
```

```
--------------------------------------------------------------------------------
Hyperparams: rf_strongly_regularized
our      | R²: 0.997 ± 0.001 | MSE: 33.867 | MAE: 2.218
sklearn  | R²: 0.995 ± 0.002 | MSE: 47.927 | MAE: 2.818
--------------------------------------------------------------------------------
Hyperparams: rf_bagging_only
our      | R²: 1.000 ± 0.000 | MSE: 1.792 | MAE: 0.712
sklearn  | R²: 1.000 ± 0.000 | MSE: 1.899 | MAE: 0.711
================================================================================
Dataset: Large dataset with high noise
================================================================================
Generated regression data with 10000 samples and 2 features.
First 5 samples of X:
 [[-0.5691482   1.59040357]
 [ 0.73487779  0.49097495]
 [ 0.20069869  1.10623156]
 [-1.07774393 -0.54427443]
 [ 0.10507597 -0.88502863]]
First 5 samples of y:
 [-56.9869474   68.99765941   2.79332014 -93.43973055   7.91252206]
--------------------------------------------------------------------------------
Hyperparams: rf_shallow
our      | R²: 0.679 ± 0.010 | MSE: 3220.896 | MAE: 41.475
sklearn  | R²: 0.696 ± 0.008 | MSE: 3042.469 | MAE: 40.270
--------------------------------------------------------------------------------
Hyperparams: rf_medium
our      | R²: 0.880 ± 0.004 | MSE: 1199.742 | MAE: 25.491
sklearn  | R²: 0.885 ± 0.005 | MSE: 1151.316 | MAE: 25.076
--------------------------------------------------------------------------------
Hyperparams: rf_deep
our      | R²: 0.953 ± 0.001 | MSE: 471.854 | MAE: 17.063
sklearn  | R²: 0.953 ± 0.002 | MSE: 466.333 | MAE: 17.008
--------------------------------------------------------------------------------
Hyperparams: rf_unrestricted
our      | R²: 0.953 ± 0.001 | MSE: 470.615 | MAE: 17.266
sklearn  | R²: 0.953 ± 0.001 | MSE: 470.539 | MAE: 17.255
--------------------------------------------------------------------------------
Hyperparams: rf_regularized_leaf
our      | R²: 0.956 ± 0.001 | MSE: 445.514 | MAE: 16.752
sklearn  | R²: 0.956 ± 0.001 | MSE: 443.462 | MAE: 16.685
--------------------------------------------------------------------------------
Hyperparams: rf_strongly_regularized
our      | R²: 0.956 ± 0.001 | MSE: 443.745 | MAE: 16.663
sklearn  | R²: 0.955 ± 0.001 | MSE: 448.253 | MAE: 16.663
--------------------------------------------------------------------------------
Hyperparams: rf_bagging_only
our      | R²: 0.953 ± 0.001 | MSE: 470.577 | MAE: 17.347
sklearn  | R²: 0.953 ± 0.001 | MSE: 469.645 | MAE: 17.351
```

```
================================================================================
Dataset: Moderate dataset with many features
================================================================================
Generated regression data with 1000 samples and 50 features.
First 5 samples of X:
 [[ 0.22210537  0.62386678 -0.16605612  0.63463021  0.75376893  1.65859202
  -0.70189811  0.18361217 -2.08195404  0.60286295  0.14837528 -1.13521912
   0.9986041  -0.49461035 -0.51024062  0.93668094 -0.399035   -0.38356556
   1.1898542  -0.5212666  -1.16138215  1.72223231  0.36539182  1.33601969
   0.78801999 -0.83774995  0.56658811  0.92829461 -0.38674766  0.6208172
   0.19797213  1.33836855  0.02365751 -2.52871576  1.75988802 -1.207332
  -0.82470391 -0.68112696  0.6151632   0.35782542 -0.7864838   0.10673978
   0.69987124 -0.03436237 -0.11283199 -0.36389558  0.10516231 -0.53678647
  -0.0572436  -0.30034055]
 [ 0.83057101 -1.71669373 -0.15501343  1.11435578 -0.82530382  0.60915961
  -0.06801749 -0.71940537 -0.05022968 -1.14378475  0.78259513 -0.10643051
   1.68403763 -0.58623557 -0.40156696 -0.59052457  0.96258821  0.79276514
  -0.16365711  0.99107132 -1.08776862 -0.43419762 -1.11047912 -1.56341347
  -0.20934306  0.19477547  1.77209275 -0.18015897  1.54042895  1.66236667
   0.17283733  0.34237509  1.7280638  -0.85290707 -1.17369964  0.55164771
  -1.65811325  0.11898642  0.20540077 -0.43031452  2.22904372 -1.44808321
  -1.14136697 -0.10875609 -0.51455358 -0.14546159  0.740684    0.85154814
  -0.72018801 -0.39993873]
 [-0.96335321  0.5437933  -1.11313988  1.21235947  0.55206656 -0.91026463
  -0.85609014  0.53390779  0.38681169  0.25747809  0.40567447 -1.38173128
  -1.34554208  2.22468763  0.15646961 -1.15238917 -0.11994209 -0.0888014
  -1.33626712 -2.46829003  0.73874622 -0.74254137  0.21518837  2.30103361
   0.6396049  -1.00437615  0.009012    1.53905609  0.68941732 -0.81846411
  -0.33594849  1.13719016 -0.84854273 -0.92413914 -0.1971291  -1.59092134
  -0.20260583  1.09189461 -1.43450227 -1.83909427  0.63691302  1.81273542
  -0.4661267   0.98463866 -0.82158042 -0.80044475 -1.55415405 -0.79595511
   0.85645506 -0.13711049]
 [ 2.1498434  -0.10467932 -0.43722006 -0.33944537  1.17827468  0.60539814
   0.62604166  0.02681872 -0.59611361 -0.325793   -0.26361207  1.89754169
  -0.40314991 -0.14655815  1.29817434  0.3420916  -0.23200561  1.83926104
  -1.72979741  2.16414981  1.01110931 -0.68284508  0.84036538  1.90220597
   2.31782164  0.93558607 -0.64628821 -0.42569971 -0.40269403 -1.27318569
   1.76253455  1.41378331 -0.04727162  1.15564512 -1.09488559  0.69360613
  -0.94454315  0.02231656  0.04946641 -2.29909535 -0.15431771  0.0542324
   0.31923321  0.39123279  0.5049892  -0.00548364 -0.58436466  0.76151403
  -0.75676622 -1.26037829]
 [-0.5372907   0.06883836 -0.49145363 -2.14575896 -0.98773148 -0.43352192
   0.31588178  0.46436161 -0.79141506 -0.20844714 -0.74615901 -0.11587905
   0.32387858  0.28785406 -0.58537497  1.2066765  -0.70211482  0.95639433
  -0.69194153 -0.20713186  0.54609752  0.32999556 -0.36007058  0.95665943
   0.22027397  1.11085552  0.40401589  0.95472592 -0.27240938 -1.51210178
   0.50589761 -0.90487118 -0.41419898 -0.42887986 -2.38069159 -1.0825947
  -0.5624822   0.92588024 -0.8635109   0.42003858 -1.82769117  0.95275321
  -1.41406661 -0.36089524  1.41201599  1.56501888  1.34935979 -0.75915137
  -0.91560975  0.18396914]]
First 5 samples of y:
 [137.83775667  -3.71055967 145.69973099  -9.68172968 -83.67934132]
--------------------------------------------------------------------------------
Hyperparams: rf_shallow
our      | R²: 0.268 ± 0.021 | MSE: 20293.013 | MAE: 112.459
sklearn  | R²: 0.305 ± 0.018 | MSE: 19282.187 | MAE: 109.292
--------------------------------------------------------------------------------
Hyperparams: rf_medium
our      | R²: 0.466 ± 0.016 | MSE: 14825.161 | MAE: 95.356
sklearn  | R²: 0.487 ± 0.019 | MSE: 14245.598 | MAE: 93.528
--------------------------------------------------------------------------------
Hyperparams: rf_deep
our      | R²: 0.644 ± 0.017 | MSE: 9892.831 | MAE: 76.654
sklearn  | R²: 0.653 ± 0.023 | MSE: 9652.582 | MAE: 75.588
--------------------------------------------------------------------------------
Hyperparams: rf_unrestricted
our      | R²: 0.683 ± 0.018 | MSE: 8813.206 | MAE: 71.746
sklearn  | R²: 0.695 ± 0.011 | MSE: 8474.058 | MAE: 70.648
--------------------------------------------------------------------------------
Hyperparams: rf_regularized_leaf
our      | R²: 0.667 ± 0.020 | MSE: 9259.127 | MAE: 73.524
sklearn  | R²: 0.661 ± 0.022 | MSE: 9424.727 | MAE: 74.527
--------------------------------------------------------------------------------
Hyperparams: rf_strongly_regularized
our      | R²: 0.631 ± 0.015 | MSE: 10252.295 | MAE: 77.665
sklearn  | R²: 0.613 ± 0.017 | MSE: 10755.422 | MAE: 79.742
--------------------------------------------------------------------------------
Hyperparams: rf_bagging_only
our      | R²: 0.881 ± 0.014 | MSE: 3289.210 | MAE: 42.954
sklearn  | R²: 0.879 ± 0.016 | MSE: 3351.223 | MAE: 43.271
```

1.4 Results

[5]:

```python
df = pd.DataFrame(results)
paired = (
    df.pivot_table(
        index=["dataset", "hyperparams"],
        columns="model",
        values=["r2_mean", "mse_mean", "mae_mean"]
    )
)

paired.columns = ["_".join(col) for col in paired.columns]
paired = paired.reset_index()
```

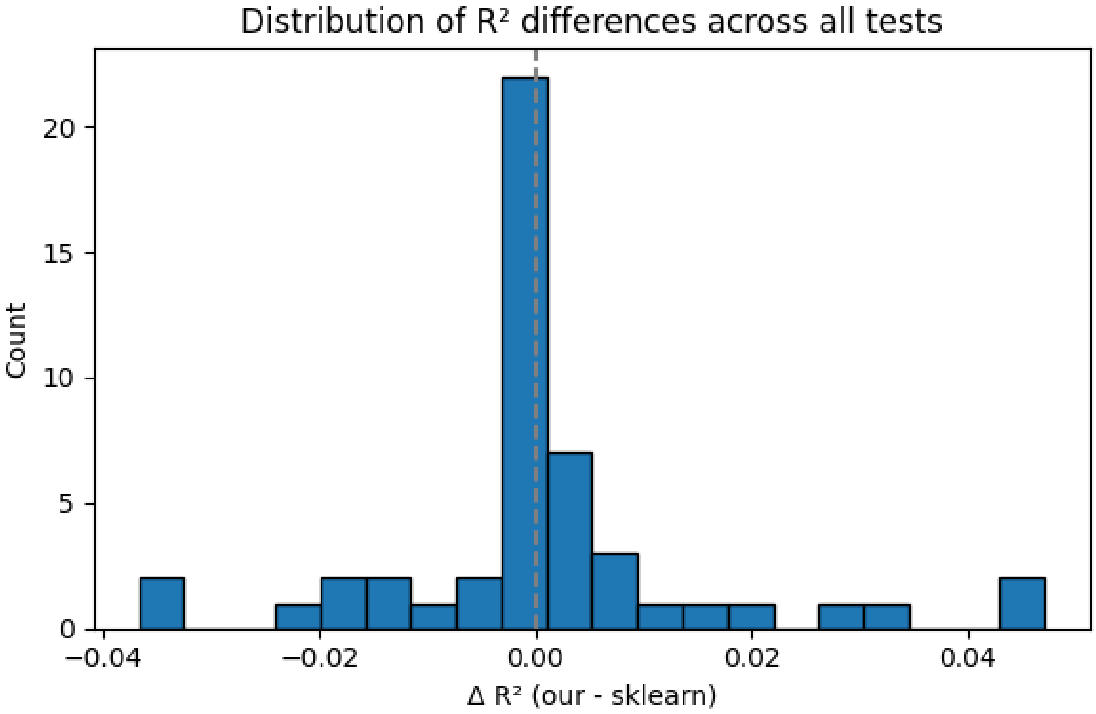
```python
paired["r2_diff"]  = paired["r2_mean_our"]  - paired["r2_mean_sklearn"]
paired["mse_diff"] = paired["mse_mean_our"] - paired["mse_mean_sklearn"]
paired["mae_diff"] = paired["mae_mean_our"] - paired["mae_mean_sklearn"]

paired.head()
```

[5]:

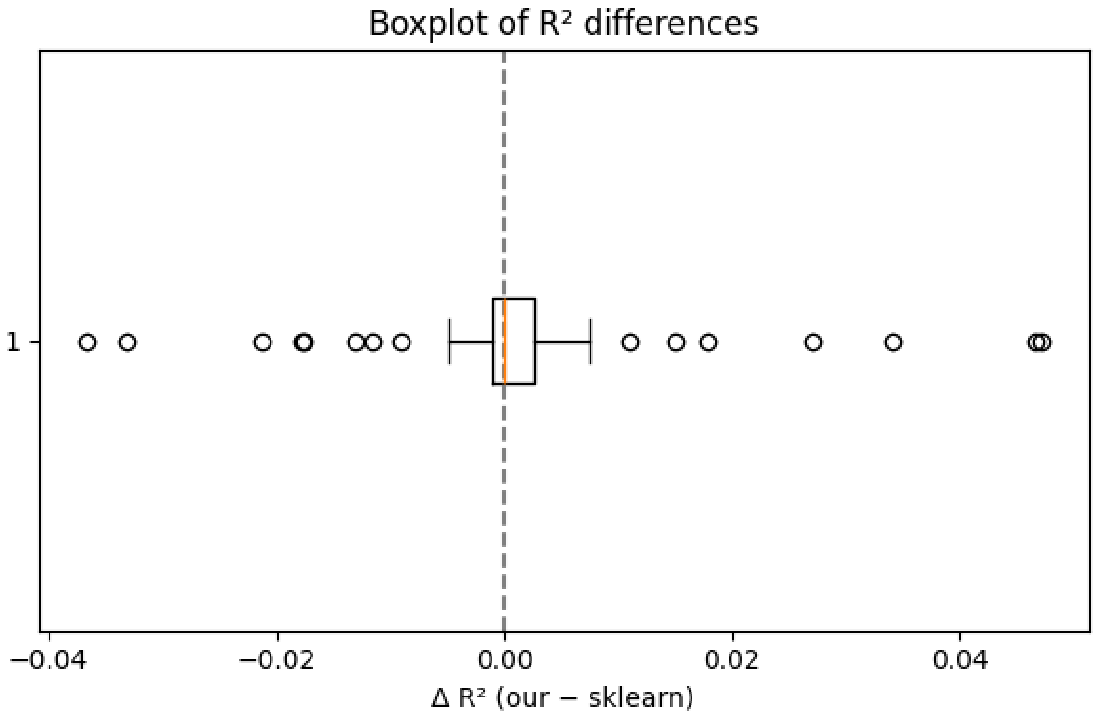
```
                                dataset          hyperparams  mae_mean_our  \
0  Basic regression data with low noise      rf_bagging_only      1.333733
1  Basic regression data with low noise              rf_deep      1.334308
2  Basic regression data with low noise            rf_medium      2.156377
3  Basic regression data with low noise  rf_regularized_leaf      2.354010
4  Basic regression data with low noise           rf_shallow      6.466155

   mae_mean_sklearn  mse_mean_our  mse_mean_sklearn  r2_mean_our  \
0          1.274271     11.849615         11.249964     0.992930
1          1.274359     11.850647         11.250346     0.992929
2          2.098668     15.910313         15.182812     0.989953
3          3.943339     25.582914         48.628196     0.984780
4          6.230456     83.495577         82.845285     0.940749

   r2_mean_sklearn   r2_diff   mse_diff  mae_diff
0         0.993262 -0.000331   0.599651  0.059462
1         0.993261 -0.000332   0.600301  0.059949
2         0.990378 -0.000425   0.727501  0.057709
3         0.969691  0.015090 -23.045282 -1.589329
4         0.940898 -0.000150   0.650291  0.235699
```

[6]:

```python
plt.figure(figsize=(6, 4))
plt.hist(paired["r2_diff"], bins=20, edgecolor="black")
plt.axvline(0, linestyle="--", color="gray")
plt.xlabel("Δ R² (our - sklearn)")
plt.ylabel("Count")
plt.title("Distribution of R² differences across all tests")
plt.tight_layout()
plt.show()
```

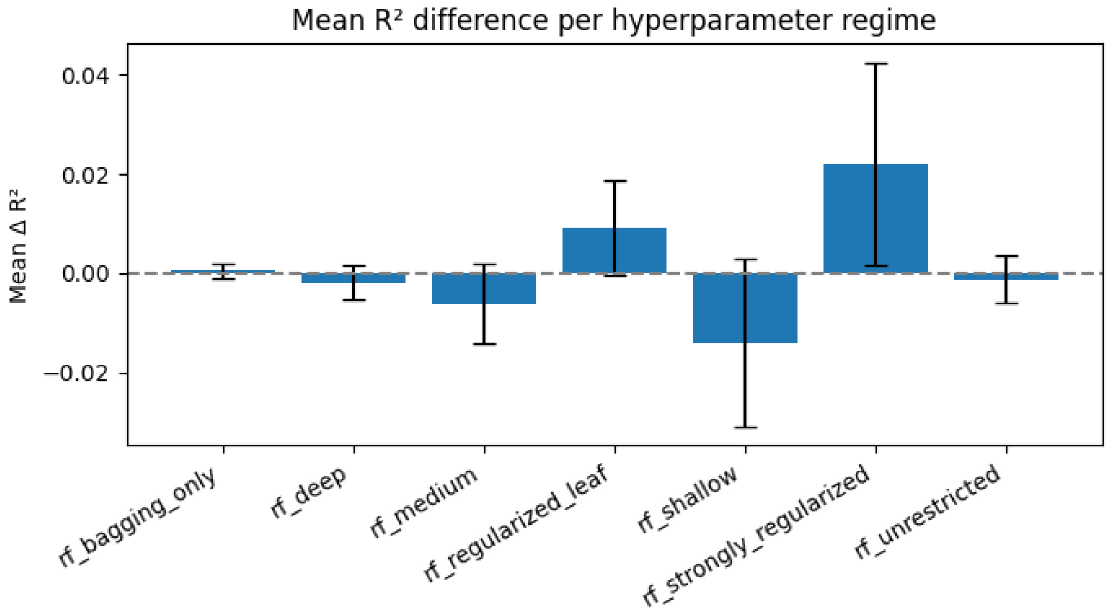


[7]:
```
plt.figure(figsize=(6, 4))
plt.boxplot(
    paired["r2_diff"],
    vert=False,
    showfliers=True
)
plt.axvline(0, linestyle="--", color="gray")
plt.xlabel("Δ R² (our - sklearn)")
plt.title("Boxplot of R² differences")
plt.tight_layout()
plt.show()
```



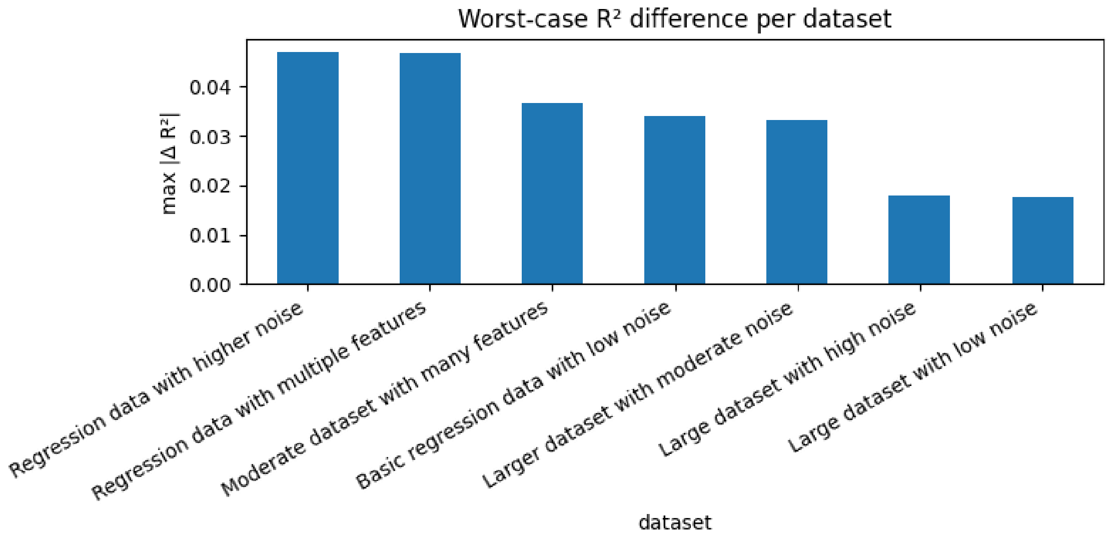
[8]:

```
by_hp = (
    paired.groupby("hyperparams")["r2_diff"]
    .agg(["mean", "std"])
    .reset_index()
)

plt.figure(figsize=(7, 4))
plt.bar(
    by_hp["hyperparams"],
    by_hp["mean"],
    yerr=by_hp["std"],
    capsize=5
)
plt.axhline(0, linestyle="--", color="gray")
plt.ylabel("Mean Δ R²")
plt.title("Mean R² difference per hyperparameter regime")
plt.xticks(rotation=30, ha="right")
plt.tight_layout()
plt.show()
```



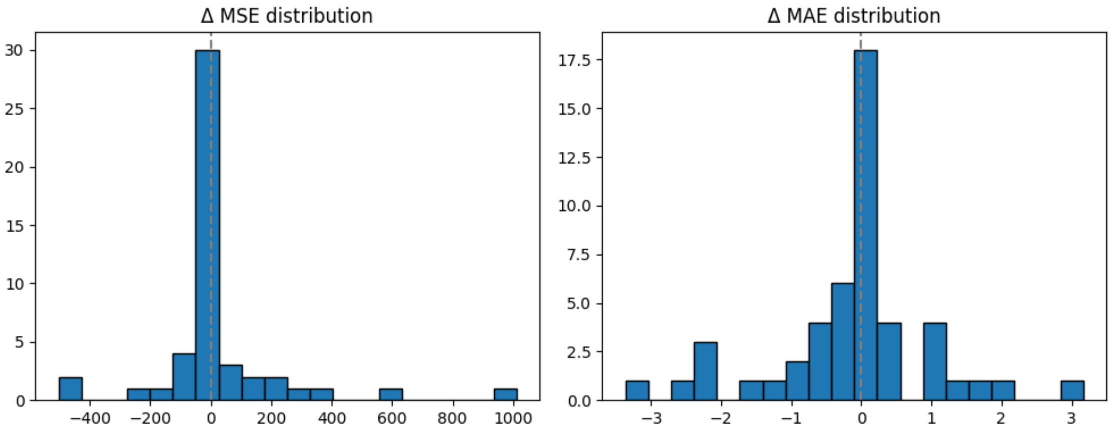
[9]:

```
worst_by_dataset = (
    paired.groupby("dataset")["r2_diff"]
    .apply(lambda x: x.abs().max())
    .sort_values(ascending=False)
)

plt.figure(figsize=(8, 4))
worst_by_dataset.plot(kind="bar")
plt.ylabel("max |Δ R²|")
plt.title("Worst-case R² difference per dataset")
plt.xticks(rotation=30, ha="right")
plt.tight_layout()
plt.show()
```



[10]:

```
fig, axes = plt.subplots(1, 2, figsize=(10, 4))

axes[0].hist(paired["mse_diff"], bins=20, edgecolor="black")
axes[0].axvline(0, linestyle="--", color="gray")
axes[0].set_title("Δ MSE distribution")

axes[1].hist(paired["mae_diff"], bins=20, edgecolor="black")
axes[1].axvline(0, linestyle="--", color="gray")
axes[1].set_title("Δ MAE distribution")

plt.tight_layout()
plt.show()
```


Δ MSE distribution
Δ MAE distribution