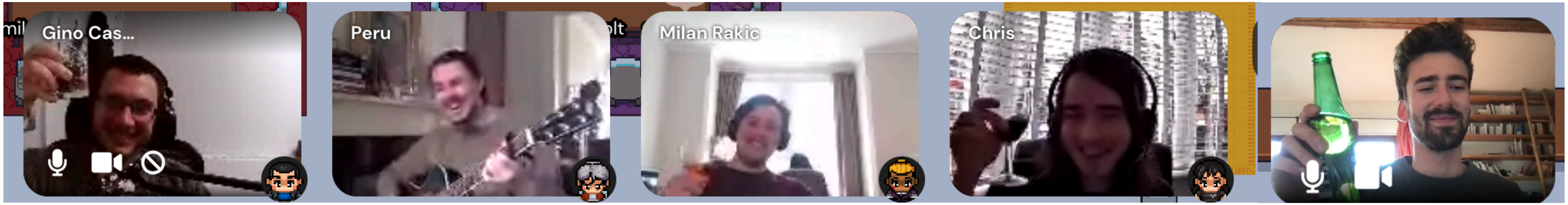
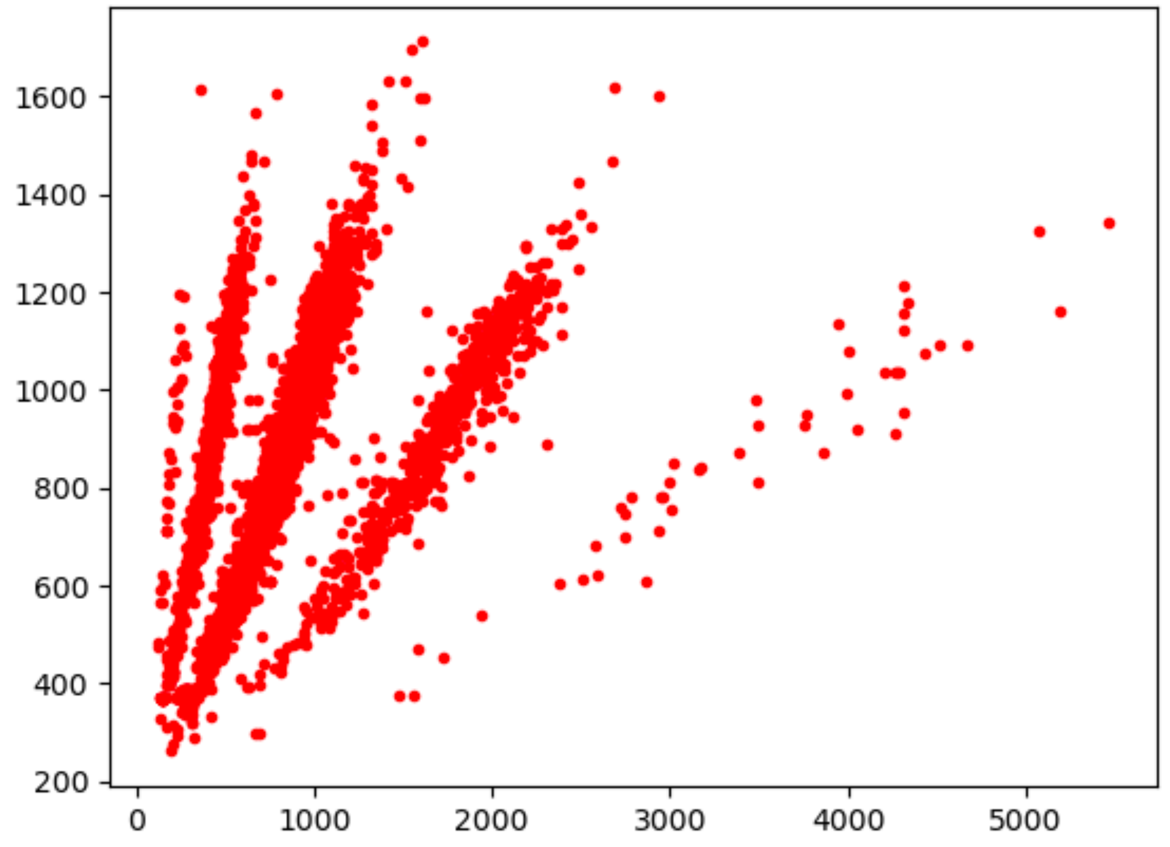


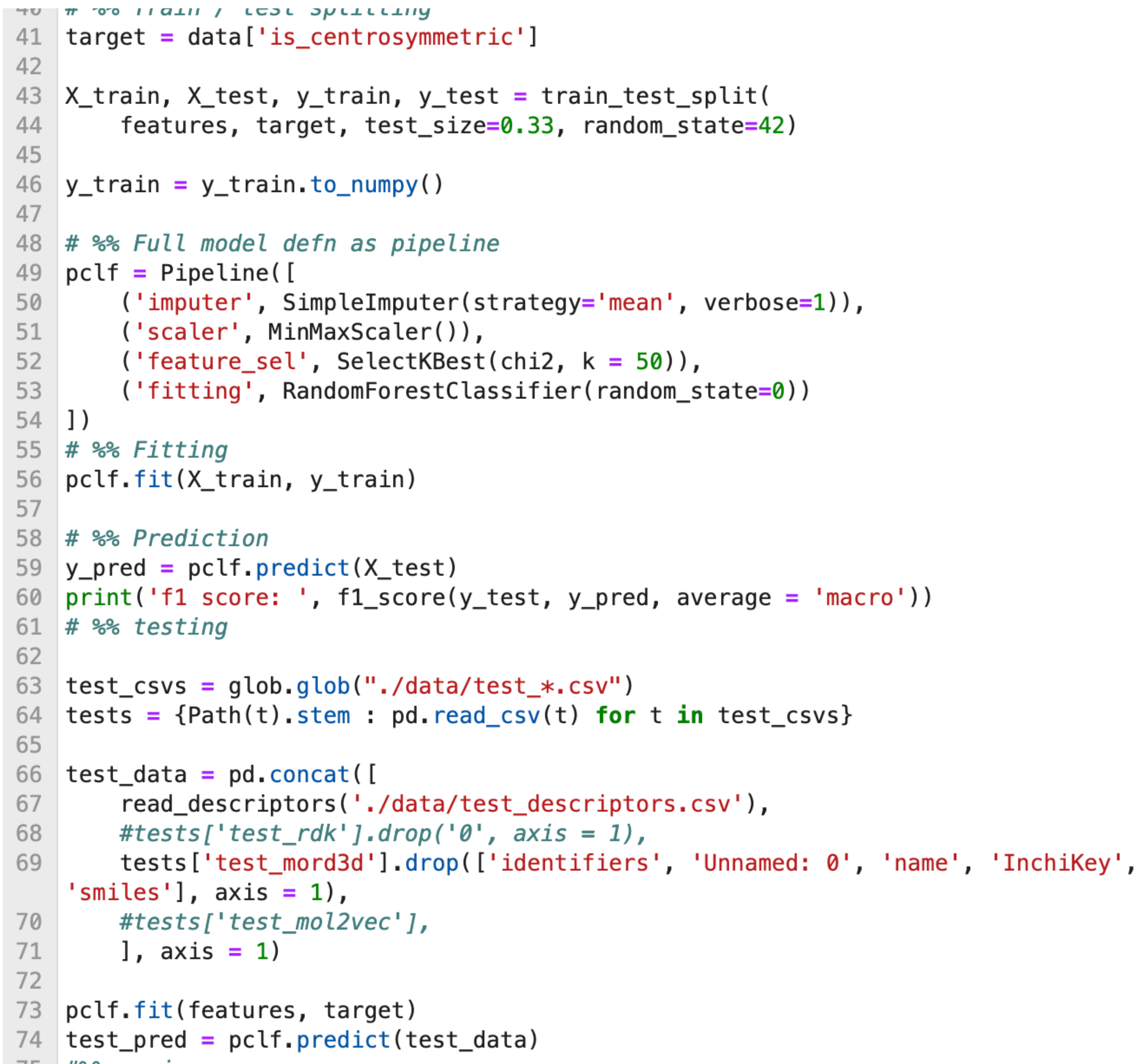
Condensed Matter Theory Group

Tom, Gino, Milan, Peru, Chris

24 March 2021

Used SKlearn standard methods.



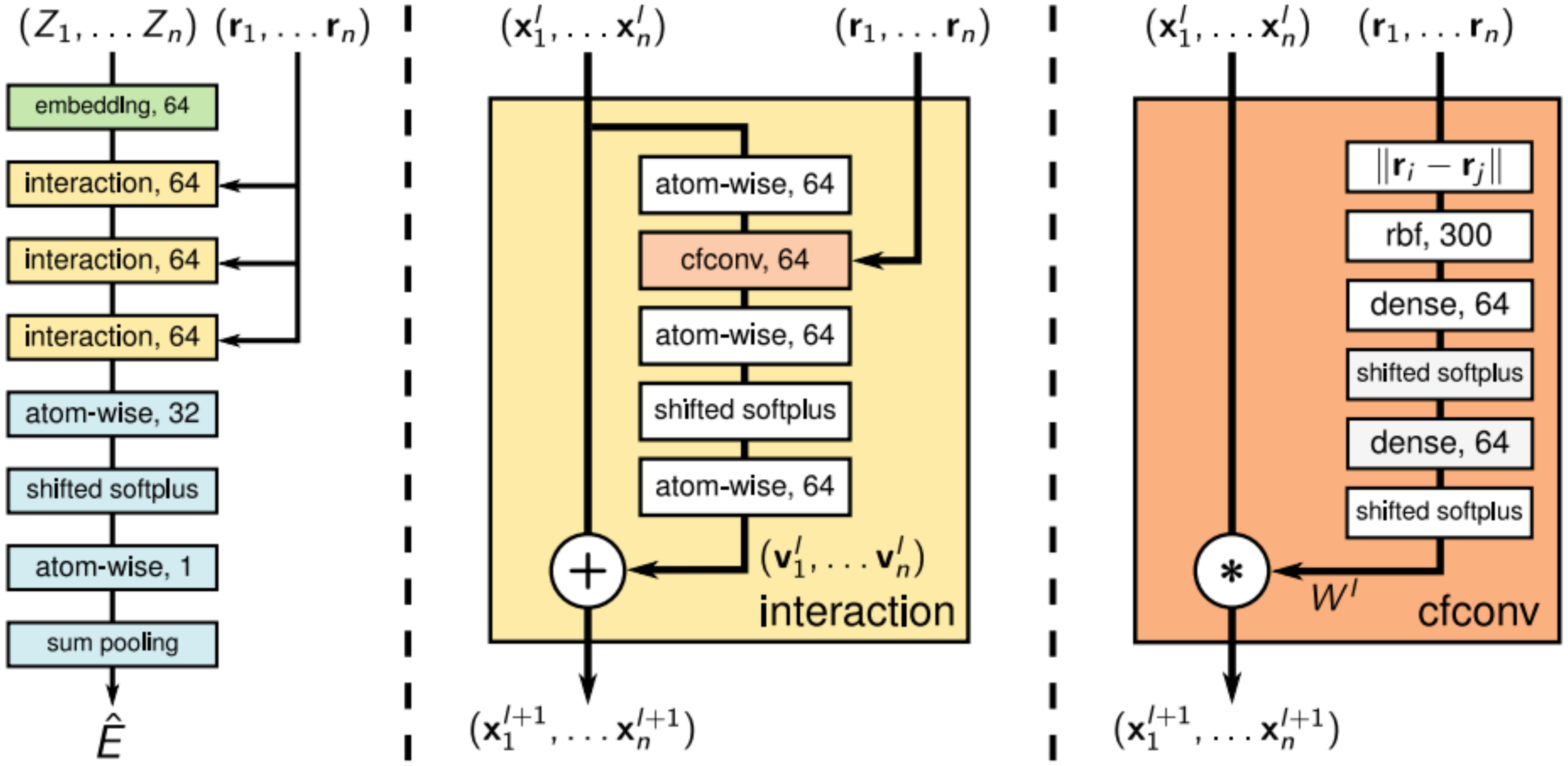


```python
# %% Train / test splitting
target = data['is_centrosymmetric']

X_train, X_test, y_train, y_test = train_test_split(
    features, target, test_size=0.33, random_state=42)

y_train = y_train.to_numpy()

# %% Full model defn as pipeline
pclf = Pipeline([
    ('imputer', SimpleImputer(strategy='mean', verbose=1)),
    ('scaler', MinMaxScaler()),
    ('feature_sel', SelectKBest(chi2, k = 50)),
    ('fitting', RandomForestClassifier(random_state=0))
])
# %% Fitting
pclf.fit(X_train, y_train)

# %% Prediction
y_pred = pclf.predict(X_test)
print('f1 score: ', f1_score(y_test, y_pred, average = 'macro'))
# %% testing

test_csvs = glob.glob("./data/test_*.csv")
tests = {Path(t).stem : pd.read_csv(t) for t in test_csvs}

test_data = pd.concat([
    read_descriptors('./data/test_descriptors.csv'),
    #tests['test_rdk'].drop('0', axis = 1),
    tests['test_mord3d'].drop(['identifiers', 'Unnamed: 0', 'name', 'InchiKey', 'smiles'], axis = 1),
    #tests['test_mol2vec'],
    ], axis = 1)

pclf.fit(features, target)
test_pred = pclf.predict(test_data)
```

- Random Forests for categorisation.
- Linear Kernal Ridge Regression on the regression.
- Used default sklearn feature scaling and imputers for cleanup.
- Refit on the full train data before doing a final prediction.
- Above: Z' in the cell volume.

Other ideas

- Mol2vec: Unsupervised Machine Learning Approach with Chemical Intuition
- SchNet: A continuous-filter convolutional neural network for modeling quantum interactions



$(Z_1, \dots Z_n)$
$(\mathbf{r}_1, \dots \mathbf{r}_n)$
embedding, 64
interaction, 64
interaction, 64
interaction, 64
atom-wise, 32
shifted softplus
atom-wise, 1
sum pooling
$\hat{E}$
$(\mathbf{x}^l_1, \dots \mathbf{x}^l_n)$
$(\mathbf{r}_1, \dots \mathbf{r}_n)$
atom-wise, 64
cfconv, 64
atom-wise, 64
shifted softplus
atom-wise, 64
$+$
$(\mathbf{v}^l_1, \dots \mathbf{v}^l_n)$
interaction
$(\mathbf{x}^{l+1}_1, \dots \mathbf{x}^{l+1}_n)$
$(\mathbf{x}^l_1, \dots \mathbf{x}^l_n)$
$(\mathbf{r}_1, \dots \mathbf{r}_n)$
$\|\mathbf{r}_i - \mathbf{r}_j\|$
rbf, 300
dense, 64
shifted softplus
dense, 64
shifted softplus
$*$
W^l
cfconv
$(\mathbf{x}^{l+1}_1, \dots \mathbf{x}^{l+1}_n)$