

```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt


from utils import *

paths = (
    '001_2/', '004_3/', '007_2/', '010_2/', '013_2/', '016_3/', '019_2/',
    '002_2/', '005_2/', '008_3/', '011_2/', '014_2/', '017_2/', '020_3/',
    '003_2/', '006_2/', '009_2/', '012_3/', '015_2/', '018_2/'
)
```

▼ Часть 1 - Кросс-валидация со всеми признаками

```
path_prefix = '20_CV/'
RMSE_train_dict, MaxAE_train_dict = get_train_RMSE_and_MaxAE(paths, path_prefix)
```

```
RMSE_train = get_dataframe_from_dict(RMSE_train_dict)
RMSE_train
```

|  | 1        | 2        | 3        | 4        | 5        | 6        |
|--|----------|----------|----------|----------|----------|----------|
| 001_2/   | 0.120919 | 0.084082 | 0.062760 | 0.051472 | 0.043619 | 0.038107 |
| 004_3/   | 0.120302 | 0.072518 | 0.056587 | 0.046938 | 0.041302 | 0.036185 |
| 007_2/   | 0.121456 | 0.084417 | 0.061506 | 0.050379 | 0.042991 | 0.037381 |
| 010_2/   | 0.093744 | 0.064102 | 0.055371 | 0.044211 | 0.038469 | 0.033682 |
| 013_2/   | 0.116201 | 0.076903 | 0.061246 | 0.050796 | 0.041229 | 0.035357 |
| 016_3/   | 0.120578 | 0.077606 | 0.057131 | 0.047842 | 0.040956 | 0.034150 |
| 019_2/   | 0.111036 | 0.060873 | 0.048292 | 0.040735 | 0.031922 | 0.026586 |
| 002_2/   | 0.122298 | 0.084100 | 0.058906 | 0.045388 | 0.034573 | 0.029597 |
| 005_2/   | 0.120458 | 0.082319 | 0.067157 | 0.049982 | 0.037691 | 0.034106 |
| 008_3/   | 0.119678 | 0.087418 | 0.066172 | 0.049741 | 0.042254 | 0.033021 |
| 011_2/   | 0.122288 | 0.083724 | 0.065022 | 0.054020 | 0.044243 | 0.035111 |
| 014_2/   | 0.110597 | 0.077155 | 0.066311 | 0.055831 | 0.042268 | 0.037741 |
| 017_2/   | 0.121850 | 0.082442 | 0.067057 | 0.054958 | 0.046875 | 0.035091 |
| 020_3/   | 0.122114 | 0.081254 | 0.066421 | 0.052049 | 0.040875 | 0.035573 |
| 003_2/   | 0.121253 | 0.084457 | 0.060592 | 0.049094 | 0.040302 | 0.034595 |
| 006_2/   | 0.118483 | 0.082420 | 0.063865 | 0.046196 | 0.040106 | 0.035099 |
| 009_2/   | 0.121782 | 0.083845 | 0.062508 | 0.048202 | 0.040885 | 0.033569 |
| 012_3/   | 0.121376 | 0.083739 | 0.067331 | 0.049657 | 0.039303 | 0.035616 |
| 015_2/   | 0.121939 | 0.089863 | 0.069525 | 0.052572 | 0.044592 | 0.039312 |
| 018_2/   | 0.121423 | 0.083482 | 0.061531 | 0.050652 | 0.040443 | 0.035017 |
| mean   | 0.118489 | 0.080336 | 0.062265 | 0.049536 | 0.040745 | 0.034745 |

```
MaxAE_train = get_dataframe_from_dict(MaxAE_train_dict)
MaxAE_train
```

|        | 1        | 2        | 3        | 4        | 5        | 6        |
|--------|----------|----------|----------|----------|----------|----------|
| 001_2/ | 0.470366 | 0.260929 | 0.146011 | 0.151498 | 0.089491 | 0.082024 |
| 004_3/ | 0.480658 | 0.240738 | 0.145557 | 0.118380 | 0.098618 | 0.097534 |
| 007_2/ | 0.456261 | 0.259297 | 0.139964 | 0.129346 | 0.098664 | 0.086051 |

|               |          |          |          |          |          |          |
|---------------|----------|----------|----------|----------|----------|----------|
| <b>010_2/</b> | 0.333910 | 0.177327 | 0.151028 | 0.098466 | 0.093490 | 0.092018 |
| <b>013_2/</b> | 0.484294 | 0.202097 | 0.158287 | 0.139703 | 0.111492 | 0.079888 |
| <b>016_3/</b> | 0.377936 | 0.236615 | 0.114585 | 0.111764 | 0.100849 | 0.083262 |
| <b>019_2/</b> | 0.543117 | 0.149160 | 0.122075 | 0.091572 | 0.077055 | 0.063245 |
| <b>002_2/</b> | 0.479974 | 0.258365 | 0.122984 | 0.149788 | 0.075315 | 0.074220 |
| <b>005_2/</b> | 0.479458 | 0.259407 | 0.241699 | 0.174367 | 0.077688 | 0.066542 |
| <b>008_3/</b> | 0.442839 | 0.325556 | 0.202026 | 0.139051 | 0.120524 | 0.077604 |
| <b>011_2/</b> | 0.478719 | 0.239221 | 0.170075 | 0.111236 | 0.112983 | 0.089412 |
| <b>014_2/</b> | 0.249682 | 0.227810 | 0.160260 | 0.168492 | 0.095471 | 0.089129 |
| <b>017_2/</b> | 0.481718 | 0.238778 | 0.160079 | 0.113472 | 0.107138 | 0.086892 |
| <b>020_3/</b> | 0.467387 | 0.253384 | 0.206111 | 0.125542 | 0.103034 | 0.091947 |
| <b>003_2/</b> | 0.492843 | 0.259101 | 0.145976 | 0.136370 | 0.106804 | 0.086323 |
| <b>006_2/</b> | 0.488641 | 0.259691 | 0.181083 | 0.106377 | 0.090629 | 0.081403 |
| <b>009_2/</b> | 0.476543 | 0.258734 | 0.142473 | 0.156945 | 0.099366 | 0.079976 |
| <b>012_3/</b> | 0.471205 | 0.260796 | 0.241637 | 0.123166 | 0.094321 | 0.089449 |

```
RMSE_valid_dict, MaxAE_valid_dict = get_validation_RMSE_and_MaxAE(paths, path_prefix, 6)
```

|               |          |          |          |          |          |          |
|---------------|----------|----------|----------|----------|----------|----------|
| <b>018_2/</b> | 0.478662 | 0.259643 | 0.139597 | 0.147450 | 0.100504 | 0.083592 |
|---------------|----------|----------|----------|----------|----------|----------|

```
RMSE_valid = get_dataframe_from_dict(RMSE_valid_dict)
RMSE_valid
```

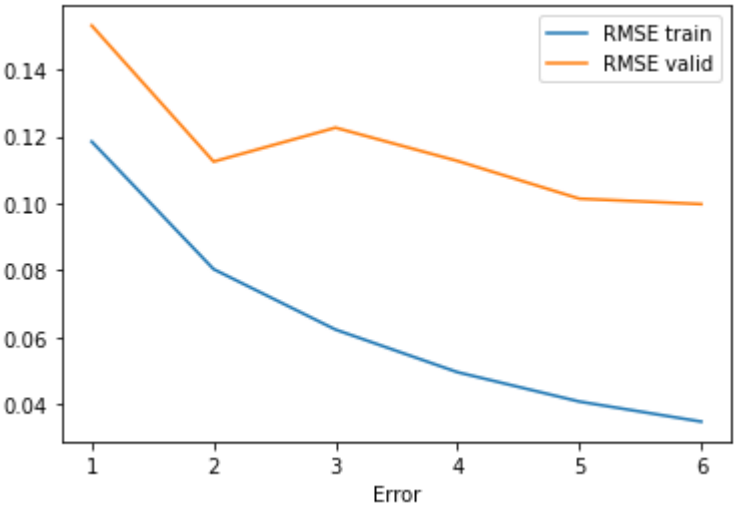
|               | 1        | 2        | 3        | 4        | 5        | 6        |
|---------------|----------|----------|----------|----------|----------|----------|
| <b>001_2/</b> | 0.092003 | 0.040390 | 0.036168 | 0.040273 | 0.030290 | 0.029487 |
| <b>004_3/</b> | 0.137363 | 0.160385 | 0.160173 | 0.084115 | 0.061299 | 0.021428 |
| <b>007_2/</b> | 0.124568 | 0.017565 | 0.063930 | 0.066297 | 0.065559 | 0.033727 |
| <b>010_2/</b> | 0.400514 | 0.301230 | 0.310669 | 0.384675 | 0.318956 | 0.448607 |
| <b>013_2/</b> | 0.179576 | 0.231349 | 0.121337 | 0.134604 | 0.090724 | 0.076035 |
| <b>016_3/</b> | 0.166245 | 0.115803 | 0.053797 | 0.061908 | 0.057575 | 0.066873 |
| <b>019_2/</b> | 0.324606 | 0.326372 | 0.303654 | 0.291341 | 0.260283 | 0.277238 |
| <b>002_2/</b> | 0.027207 | 0.038603 | 0.107855 | 0.095515 | 0.104767 | 0.104082 |
| <b>005_2/</b> | 0.104949 | 0.091456 | 0.073366 | 0.085884 | 0.082676 | 0.086615 |
| <b>008_3/</b> | 0.206787 | 0.112284 | 0.118516 | 0.102944 | 0.080266 | 0.087564 |
| <b>011_2/</b> | 0.027507 | 0.072718 | 0.068899 | 0.061982 | 0.060354 | 0.056136 |
| <b>014_2/</b> | 0.327023 | 0.182551 | 0.421978 | 0.325447 | 0.318953 | 0.238196 |
| <b>017_2/</b> | 0.055732 | 0.089650 | 0.031684 | 0.057760 | 0.047506 | 0.059262 |
| <b>020_3/</b> | 0.128781 | 0.126168 | 0.122575 | 0.042509 | 0.078785 | 0.081844 |
| <b>003_2/</b> | 0.144555 | 0.012126 | 0.090271 | 0.036047 | 0.022904 | 0.016274 |
| <b>006_2/</b> | 0.213099 | 0.104778 | 0.129757 | 0.119603 | 0.117107 | 0.037987 |
| <b>009_2/</b> | 0.058849 | 0.049163 | 0.043738 | 0.074568 | 0.048140 | 0.045014 |
| <b>012_3/</b> | 0.108242 | 0.074996 | 0.080027 | 0.080966 | 0.084133 | 0.088916 |
| <b>015_2/</b> | 0.161935 | 0.040459 | 0.045680 | 0.048031 | 0.033079 | 0.093773 |
| <b>018_2/</b> | 0.074277 | 0.062060 | 0.068841 | 0.059711 | 0.065005 | 0.046975 |
| <b>mean</b>   | 0.153191 | 0.112505 | 0.122646 | 0.112709 | 0.101418 | 0.099802 |

```
MaxAE_valid = get_dataframe_from_dict(MaxAE_valid_dict)
MaxAE_valid
```

|               | 1        | 2        | 3        | 4        | 5        | 6        |
|---------------|----------|----------|----------|----------|----------|----------|
| <b>001_2/</b> | 0.129585 | 0.056734 | 0.040884 | 0.056376 | 0.042403 | 0.041693 |
| <b>004_3/</b> | 0.194201 | 0.226763 | 0.226476 | 0.118951 | 0.083138 | 0.030302 |
| <b>007_2/</b> | 0.141718 | 0.019922 | 0.088442 | 0.093488 | 0.091054 | 0.044687 |

|               |          |          |          |          |          |          |
|---------------|----------|----------|----------|----------|----------|----------|
| <b>010_2/</b> | 0.563148 | 0.423407 | 0.435531 | 0.538877 | 0.447280 | 0.631848 |
| <b>013_2/</b> | 0.239864 | 0.284315 | 0.167425 | 0.187432 | 0.098874 | 0.080031 |
| <b>016_3/</b> | 0.166371 | 0.135033 | 0.058642 | 0.078068 | 0.061532 | 0.092577 |
| <b>019_2/</b> | 0.413371 | 0.438276 | 0.428567 | 0.408066 | 0.364640 | 0.390151 |
| <b>002_2/</b> | 0.038237 | 0.047065 | 0.148475 | 0.135023 | 0.134191 | 0.132077 |
| <b>005_2/</b> | 0.109724 | 0.097356 | 0.079006 | 0.091148 | 0.087919 | 0.090997 |
| <b>008_3/</b> | 0.256810 | 0.157288 | 0.167236 | 0.125415 | 0.084766 | 0.093784 |
| <b>011_2/</b> | 0.038404 | 0.089192 | 0.095212 | 0.086226 | 0.085327 | 0.077205 |
| <b>014_2/</b> | 0.450734 | 0.251948 | 0.594937 | 0.460252 | 0.450561 | 0.336354 |
| <b>017_2/</b> | 0.055956 | 0.109053 | 0.033814 | 0.081614 | 0.060583 | 0.078181 |
| <b>020_3/</b> | 0.163529 | 0.167385 | 0.161788 | 0.057574 | 0.107337 | 0.098911 |
| <b>003_2/</b> | 0.185291 | 0.014690 | 0.127465 | 0.048906 | 0.032336 | 0.022855 |
| <b>006_2/</b> | 0.296492 | 0.148154 | 0.180538 | 0.169029 | 0.165573 | 0.045493 |
| <b>009_2/</b> | 0.066699 | 0.069032 | 0.054401 | 0.092183 | 0.067110 | 0.046685 |
| <b>012_3/</b> | 0.124719 | 0.087576 | 0.112680 | 0.113441 | 0.094622 | 0.101703 |
| <b>015_2/</b> | 0.206749 | 0.051292 | 0.060911 | 0.066041 | 0.039567 | 0.132375 |
| <b>018_2/</b> | 0.101082 | 0.082744 | 0.083475 | 0.067910 | 0.068849 | 0.061849 |
| <b>mean</b>   | 0.197134 | 0.147861 | 0.167295 | 0.153801 | 0.133383 | 0.131488 |

```
plt.plot(RMSE_train.columns, RMSE_train.loc['mean'], label='RMSE train')
plt.plot(RMSE_valid.columns, RMSE_valid.loc['mean'], label='RMSE valid')
plt.legend()
plt.xlabel('Model complexity')
plt.xlabel('Error')
plt.show()
```



По рисунку видно, что не наступает переобученности. Поэтому лучшим будем считать размерность дескриптора с минимальной ошибкой валидации.

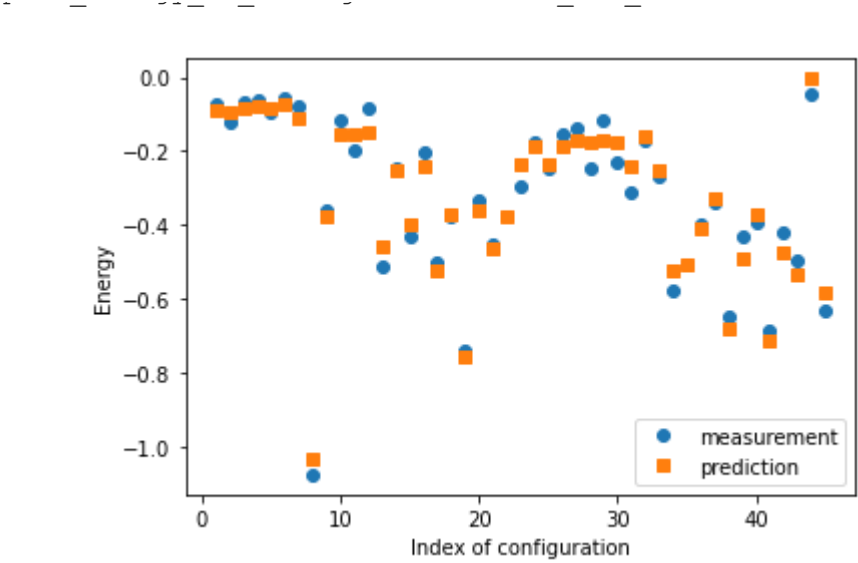
```
print(f"Минимальная ошибка RMSE валидации при размерности {RMSE_valid.columns[np.argmin(RMSE_valid.loc['mean'])]}")
print(f"Минимальная ошибка MaxAE валидации при размерности {MaxAE_valid.columns[np.argmin(MaxAE_valid.loc['mean'])]}")
```

```
Минимальная ошибка RMSE валидации при размерности 6
Минимальная ошибка MaxAE валидации при размерности 6
```

По полученным результатам становится ясно, что наилучшая размерность дескриптора – 6. Поэтому далее обучаем на всей базе с этой размерностью.

▼ Часть 2 - Обучение на полной выборке с наилучшей размерностью

```
path_prefix = 'full_set/desc_dat/'
full_set_desc = get_full_set_desc_dat(path_prefix, 6)
plot energy vs configuration(full set desc)
```



```
RMSE, MaxAE = get_full_set_desc_dat_errors(full_set_desc)
print('Обучение на полной выборке с дескриптором бго уровня:')
print(f'RMSE: {RMSE}, MaxAE: {MaxAE}')
```

Обучение на полной выборке с дескриптором бго уровня:  
RMSE: 0.036351709578525106, MaxAE: 0.064009999999999998

### ▼ Часть 3 - Кросс-валидация с одним признаком

```
path_prefix = '20_CV_only_DS/'
RMSE_train_dict, MaxAE_train_dict = get_train_RMSE_and_MaxAE(paths, path_prefix)
```

```
RMSE_train = get_dataframe_from_dict(RMSE_train_dict)
RMSE_train
```

|        | 1        | 2        | 3        | 4        | 5        | 6        |
|--------|----------|----------|----------|----------|----------|----------|
| 001_2/ | 0.210326 | 0.208484 | 0.206862 | 0.206306 | 0.204986 | 0.203679 |
| 004_3/ | 0.212150 | 0.210588 | 0.208302 | 0.208113 | 0.207294 | 0.206299 |
| 007_2/ | 0.209469 | 0.208005 | 0.205716 | 0.205549 | 0.204794 | 0.203562 |
| 010_2/ | 0.175210 | 0.174597 | 0.172806 | 0.171628 | 0.167749 | 0.152045 |
| 013_2/ | 0.199498 | 0.196629 | 0.194902 | 0.194380 | 0.192366 | 0.191997 |
| 016_3/ | 0.213525 | 0.212005 | 0.209983 | 0.208699 | 0.208580 | 0.194747 |
| 019_2/ | 0.199651 | 0.196734 | 0.194642 | 0.194459 | 0.192717 | 0.182456 |
| 002_2/ | 0.208683 | 0.208266 | 0.204604 | 0.204554 | 0.203889 | 0.190816 |
| 005_2/ | 0.203867 | 0.196712 | 0.196492 | 0.196200 | 0.195443 | 0.179028 |
| 008_3/ | 0.205234 | 0.204877 | 0.203684 | 0.200853 | 0.198900 | 0.197866 |
| 011_2/ | 0.211410 | 0.209868 | 0.207934 | 0.207549 | 0.206585 | 0.191100 |
| 014_2/ | 0.208172 | 0.207670 | 0.203870 | 0.203830 | 0.203341 | 0.201952 |
| 017_2/ | 0.208284 | 0.206722 | 0.204109 | 0.204025 | 0.203149 | 0.191932 |
| 020_3/ | 0.211190 | 0.210093 | 0.209219 | 0.209025 | 0.208487 | 0.193880 |
| 003_2/ | 0.206606 | 0.205169 | 0.203240 | 0.202792 | 0.201632 | 0.201033 |
| 006_2/ | 0.205538 | 0.204582 | 0.200141 | 0.199828 | 0.199092 | 0.189782 |
| 009_2/ | 0.211357 | 0.209590 | 0.207755 | 0.207354 | 0.206376 | 0.205156 |
| 012_3/ | 0.212952 | 0.211333 | 0.209251 | 0.208952 | 0.207970 | 0.192981 |
| 015_2/ | 0.209180 | 0.206958 | 0.206075 | 0.205992 | 0.205121 | 0.190910 |
| 018_2/ | 0.209294 | 0.208019 | 0.206112 | 0.205678 | 0.204903 | 0.188187 |
| mean   | 0.206580 | 0.204845 | 0.202785 | 0.202288 | 0.201169 | 0.192470 |

```
MaxAE_train = get_dataframe_from_dict(MaxAE_train_dict)
MaxAE_train
```

|        | 1        | 2        | 3        | 4        | 5        | 6        |
|--------|----------|----------|----------|----------|----------|----------|
| 001_2/ | 0.718878 | 0.669424 | 0.700504 | 0.710916 | 0.710606 | 0.727879 |

|        |          |          |          |          |          |          |
|--------|----------|----------|----------|----------|----------|----------|
| 004_3/ | 0.716794 | 0.671626 | 0.707440 | 0.713846 | 0.713112 | 0.727250 |
| 007_2/ | 0.715309 | 0.671378 | 0.707151 | 0.713352 | 0.712861 | 0.728261 |
| 010_2/ | 0.455497 | 0.454087 | 0.472027 | 0.465658 | 0.471678 | 0.468851 |
| 013_2/ | 0.727774 | 0.667851 | 0.698551 | 0.708775 | 0.708651 | 0.717638 |
| 016_3/ | 0.717486 | 0.666744 | 0.696612 | 0.716605 | 0.718183 | 0.680130 |
| 019_2/ | 0.724577 | 0.657333 | 0.695786 | 0.705938 | 0.713022 | 0.701824 |
| 002_2/ | 0.713976 | 0.740064 | 0.710380 | 0.713801 | 0.713346 | 0.694205 |
| 005_2/ | 0.701360 | 0.590963 | 0.604895 | 0.594241 | 0.601367 | 0.572830 |
| 008_3/ | 0.704580 | 0.681043 | 0.694335 | 0.714958 | 0.725132 | 0.744581 |
| 011_2/ | 0.715085 | 0.669754 | 0.702531 | 0.712158 | 0.711638 | 0.690672 |
| 014_2/ | 0.715894 | 0.744670 | 0.714681 | 0.711925 | 0.715630 | 0.732297 |
| 017_2/ | 0.715628 | 0.670341 | 0.708093 | 0.712549 | 0.711955 | 0.694614 |
| 020_3/ | 0.726197 | 0.686968 | 0.706219 | 0.712931 | 0.712216 | 0.687911 |
| 003_2/ | 0.709558 | 0.666228 | 0.699096 | 0.708971 | 0.708115 | 0.719026 |
| 006_2/ | 0.708033 | 0.747269 | 0.715194 | 0.707416 | 0.712313 | 0.696181 |
| 009_2/ | 0.719206 | 0.670711 | 0.702920 | 0.712518 | 0.711954 | 0.727303 |
| 012_3/ | 0.716066 | 0.669910 | 0.704513 | 0.712382 | 0.711337 | 0.695135 |
| 015_2/ | 0.718855 | 0.657486 | 0.678706 | 0.683810 | 0.683046 | 0.660344 |
| 018_2/ | 0.712339 | 0.671310 | 0.703488 | 0.713837 | 0.713345 | 0.692787 |
| mean   | 0.702655 | 0.666258 | 0.686156 | 0.692329 | 0.693975 | 0.687986 |

```
RMSE_valid_dict, MaxAE_valid_dict = get_validation_RMSE_and_MaxAE(paths, '20_CV_only_DS/', 6)
```

```
RMSE_valid = get_dataframe_from_dict(RMSE_valid_dict)
RMSE_valid
```

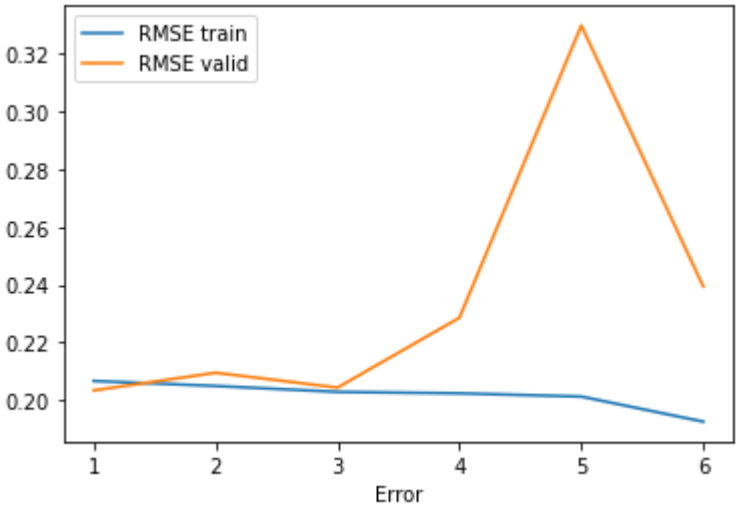
|        | 1        | 2        | 3        | 4        | 5        | 6        |
|--------|----------|----------|----------|----------|----------|----------|
| 001_2/ | 0.120756 | 0.128180 | 0.109408 | 0.125921 | 0.135302 | 0.145654 |
| 004_3/ | 0.141312 | 0.137840 | 0.144400 | 0.140915 | 0.134464 | 0.145671 |

```
MaxAE_valid = get_dataframe_from_dict(MaxAE_valid_dict)
MaxAE_valid
```

MaxAE\_valid

|        | 1        | 2        | 3        | 4        | 5        | 6        |
|--------|----------|----------|----------|----------|----------|----------|
| 001_2/ | 0.157495 | 0.168671 | 0.146347 | 0.164780 | 0.171443 | 0.200090 |
| 004_3/ | 0.143722 | 0.144889 | 0.161348 | 0.153117 | 0.146555 | 0.148936 |
| 007_2/ | 0.197727 | 0.184424 | 0.205017 | 0.196232 | 0.186478 | 0.185522 |
| 010_2/ | 0.851269 | 0.865804 | 0.817998 | 0.834691 | 0.840444 | 0.882530 |
| 013_2/ | 0.424184 | 0.442936 | 0.443871 | 0.442618 | 0.457187 | 0.443723 |
| 016_3/ | 0.113257 | 0.082900 | 0.132351 | 0.356508 | 0.544139 | 0.537909 |
| 019_2/ | 0.473019 | 0.491689 | 0.492974 | 0.491544 | 0.507511 | 0.449302 |
| 002_2/ | 0.239803 | 0.250595 | 0.247952 | 0.243699 | 0.235712 | 0.128549 |
| 005_2/ | 0.290734 | 0.401064 | 0.387168 | 0.397782 | 0.390504 | 0.459101 |
| 008_3/ | 0.350937 | 0.356447 | 0.333142 | 0.664264 | 3.487801 | 0.930025 |
| 011_2/ | 0.089865 | 0.083785 | 0.068974 | 0.080486 | 0.085242 | 0.171589 |
| 014_2/ | 0.224031 | 0.235756 | 0.233166 | 0.235711 | 0.223274 | 0.218550 |
| 017_2/ | 0.218960 | 0.205019 | 0.225805 | 0.220691 | 0.209384 | 0.181519 |
| 020_3/ | 0.237633 | 0.213336 | 0.160740 | 0.153450 | 0.114663 | 0.067698 |
| 003_2/ | 0.261961 | 0.262202 | 0.243410 | 0.258082 | 0.273969 | 0.254754 |
| 006_2/ | 0.248453 | 0.271882 | 0.299843 | 0.316364 | 0.314719 | 0.209666 |
| 009_2/ | 0.102259 | 0.113459 | 0.088448 | 0.103777 | 0.107409 | 0.124556 |
| 012_3/ | 0.094955 | 0.094794 | 0.115546 | 0.102780 | 0.087856 | 0.079838 |
| 015_2/ | 0.161305 | 0.222654 | 0.201339 | 0.196656 | 0.197204 | 0.088195 |
| 018_2/ | 0.177390 | 0.165163 | 0.161079 | 0.164426 | 0.156812 | 0.211620 |
| mean   | 0.252948 | 0.262873 | 0.258326 | 0.288883 | 0.436915 | 0.298684 |

```
plt.plot(RMSE_train.columns, RMSE_train.loc['mean'], label='RMSE train')
plt.plot(RMSE_valid.columns, RMSE_valid.loc['mean'], label='RMSE valid')
plt.legend()
plt.xlabel('Model complexity')
plt.xlabel('Error')
plt.show()
```



По рисунку видно, что не наступает переобученности. Поэтому лучшим будем считать размерность дескриптора с минимальной ошибкой валидации.

```
print(f"Минимальная ошибка RMSE валидации при размерности {RMSE_valid.columns[np.argmin(RMSE_valid.loc['mean'])]}")
print(f"Минимальная ошибка MaxAE валидации при размерности {MaxAE_valid.columns[np.argmin(MaxAE_valid.loc['mean'])]}")
```

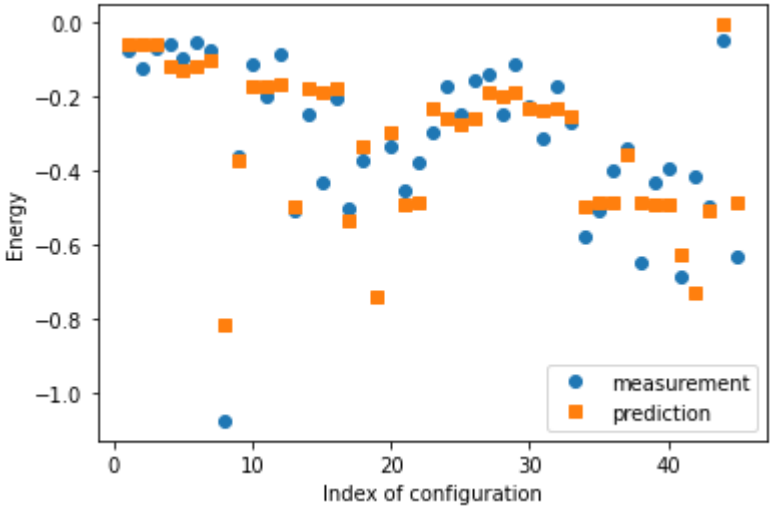
Минимальная ошибка RMSE валидации при размерности 1  
Минимальная ошибка MaxAE валидации при размерности 1

Хоть минимум ошибки валидации достигается при размерности 1, мы будем использовать далее размерность 3, т.к. это лучшая размерность до переобучения.

➤ Часть 4 - Обвчение на полной выборке с одним признаком

... DS - это не единственный значимый признак.

```
full_set_desc_only_DS = get_full_set_desc_dat('full_set_only_DS/desc_dat/', 6)
plot_energy_vs_configuration(full_set_desc_only_DS)
```



```
RMSE, MaxAE = get_full_set_desc_dat_errors(full_set_desc_only_DS)
print('Обучение на полной выборке с дескриптором 6го уровня:')
print(f'RMSE: {RMSE}, MaxAE: {MaxAE}')
```

Обучение на полной выборке с дескриптором 6го уровня:  
RMSE: 0.09458696504880278, MaxAE: 0.31166

## Итоги

Итак, ошибка предсказания на выборке с одним признаком больше, чем при использовании всех параметров. Это выглядит разумно, т.к. выбранный нами признак DS - это не единственный значимый признак. Это можно понять по анализу тренировочного набора.