Fitting drug response curves with sigmoid function

In [4]:

```
import pandas as pd
import numpy as np
from tqdm import tqdm
import warnings
warnings.filterwarnings("ignore")
import os, sys
sys.path.insert(1, os.path.relpath("functions"))
from fitting import *
from plotting import *
R2_limit = 0.99

_FOLDER = "results/"
_FOLDER_2 = "database/"
```

Fitting data

In [5]:

```
df_auc = pd.read_csv(_F0LDER+"filt_auc_02.csv")
df_1234 = pd.read_csv(_F0LDER+"filt_1234_02.csv")
drug_curves = pd.read_csv(_F0LDER_2+"normalised_dose_response_data.csv")
conc_columns= ["fd_num_"+str(i) for i in range(10)]
response_norm = ['norm_cells_'+str(i) for i in range(10)]
df_auc.shape, df_1234.shape, drug_curves.shape
```

Out [5]:

```
((14084, 31), (2108, 30), (225384, 44))
```

In [6]:

```
functions = [
    "fsigmoid",
    "sigmoid_2_param",
    "sigmoid_3_param",
    "sigmoid_4_param",
    "logistic_4_param",
    "I14_4_param",
    "I14R_4_param",
    "logLogist_3_param"
]
```

In [7]:

```
def compare_fitting(df_raw, fitting_functions_list):
    df_results = pd.DataFrame(columns= ["R2>0.9", "R2>0.95", "R2>0.99"])
    for fitting_function in fitting_functions_list:
       df = df_raw.copy()
        print("₩n", fitting_function)
        r2, fit_param = fitting_column(df, df.index, x_columns=conc_columns, y_columns= response)
                              fitting_function = fitting_function, default_param=True)
        df[fitting_function+"_r2"] = r2
        df[fitting_function] = fit_param
         df= df[df[fitting_function+"_r2"]>0]
       print("R2>0:", df.shape)
       print("R2>", R2_limit, df[df[fitting_function+"_r2"]>R2_limit].shape[0])
        df_results.loc[fitting_function, "R2>0.9"] = df[df[fitting_function+"_r2"]>0.9].shape[0]
        df_results.loc[fitting_function, "R2>0.95"] = df[df[fitting_function+"_r2"]>0.95].shape[(
        df_results.loc[fitting_function, "R2>0.99"] = df[df[fitting_function+"_r2"]>0.99].shape[(
       print("Number of samples with fitting <0.1:", df[df[fitting_function+"_r2"]<0.1].shape[0]
        print("")
    return df, df_results
```

```
In [8]:
```

```
%%time
df, df_results = compare_fitting(df_auc, functions)
df_results.to_csv(_FOLDER+"fit_auc_02_compare.csv", index=False)
df.to_csv(_FOLDER+"filt_auc_02_fit.csv", index=False)
fsigmoid
100%
14084/14084 [00:15<00:00, 907.74it/s]
<function fsigmoid at 0x0000020C7F7C8220>
R2>0: (14084, 33)
R2> 0.99 1423
Number of samples with fitting <0.1: 26
sigmoid_2_param
100%
14084/14084 [00:13<00:00, 1018.07it/s]
<function sigmoid_2_param at 0x0000020C7F7C85E0>
R2>0: (14084, 33)
R2> 0.99 1409
Number of samples with fitting <0.1: 12
sigmoid_3_param
100%
14084/14084 [00:23<00:00, 593.25it/s]
<function sigmoid_3_param at 0x0000020C7F7C8900>
R2>0: (14084, 33)
R2> 0.99 2507
Number of samples with fitting <0.1: 41
sigmoid_4_param
14084/14084 [02:08<00:00, 109.19it/s]
<function sigmoid_4_param at 0x0000020C7F7C89A0>
R2>0: (14084, 33)
R2> 0.99 3062
Number of samples with fitting <0.1: 4755
 logistic_4_param
14084/14084 [01:38<00:00, 142.34it/s]
```

<function logistic_4_param at 0x0000020C7F7C8B80>

R2>0: (14084, 33) R2> 0.99 3072

Number of samples with fitting <0.1: 1758

114_4_param

100%

| 14084/14084 [02:21<00:00, 99.43it/s]

<function II4_4_param at 0x0000020C7F7C8A40>

R2>0: (14084, 33) R2> 0.99 3075

Number of samples with fitting <0.1: 1813

II4R_4_param

100%

| 14084/14084 [02:44<00:00, 85.70it/s]

<function II4R_4_param at 0x0000020C7F7C8AE0>

R2>0: (14084, 33) R2> 0.99 3014

Number of samples with fitting <0.1: 2514

logLogist_3_param

100%

14084/14084 [01:09<00:00, 201.52it/s]

<function logLogist_3_param at 0x0000020C7F7C8C20>

R2>0: (14084, 33) R2> 0.99 2914

Number of samples with fitting <0.1: 458

CPU times: total: 11min 1s

Wall time: 10min 58s

```
In [9]:
```

```
%%time
df, df_results_2 = compare_fitting(df_1234, functions)
df_results_2.to_csv(_FOLDER+"fit_1234_compare.csv", index=False)
df.to_csv(_FOLDER+"filt_1234_fit.csv", index=False)
fsigmoid
100%
2108/2108 [00:01<00:00, 1204.35it/s]
<function fsigmoid at 0x0000020C7F7C8220>
R2>0: (2108, 32)
R2> 0.99 921
Number of samples with fitting <0.1: 0
sigmoid_2_param
100%
2108/2108 [00:01<00:00, 1279.44it/s]
<function sigmoid_2_param at 0x0000020C7F7C85E0>
R2>0: (2108, 32)
R2> 0.99 921
Number of samples with fitting <0.1: 0
sigmoid_3_param
100%
■| 2108/2108 [00:02<00:00, 756.22it/s]
<function sigmoid_3_param at 0x0000020C7F7C8900>
R2>0: (2108, 32)
R2> 0.99 1220
Number of samples with fitting <0.1: 0
sigmoid_4_param
100%
1 2108/2108 [00:03<00:00, 666.34it/s]
<function sigmoid_4_param at 0x0000020C7F7C89A0>
R2>0: (2108, 32)
R2> 0.99 1527
Number of samples with fitting <0.1:63
 logistic_4_param
■| 2108/2108 [00:02<00:00, 704.55it/s]
```

<function logistic_4_param at 0x0000020C7F7C8B80>

R2>0: (2108, 32) R2> 0.99 1473

Number of samples with fitting <0.1: 0

114_4_param

100%

| 2108/2108 [00:03<00:00, 541.54it/s]

<function II4_4_param at 0x0000020C7F7C8A40>

R2>0: (2108, 32) R2> 0.99 1473

Number of samples with fitting <0.1: 0

II4R_4_param

100%

| 2108/2108 [00:04<00:00, 452.60it/s]

<function II4R_4_param at 0x0000020C7F7C8AE0>

R2>0: (2108, 32) R2> 0.99 1473

Number of samples with fitting <0.1: 19

logLogist_3_param

100%

2108/2108 [00:03<00:00, 555.64it/s]

<function logLogist_3_param at 0x0000020C7F7C8C20>

R2>0: (2108, 32) R2> 0.99 1465

Number of samples with fitting <0.1: 0

CPU times: total: 25.2 s

Wall time: 25 s

In [10]:

```
%%time
df, df_results_3 = compare_fitting(drug_curves, functions)
df_results_3.to_csv(_FOLDER+"fit_no_filt_compare.csv", index=False)
df.to_csv(_FOLDER+"filt_fit.csv", index=False)
fsigmoid
100%
25384/225384 [04:48<00:00, 781.67it/s]
<function fsigmoid at 0x0000020C7F7C8220>
R2>0: (225384, 46)
R2> 0.99 6638
Number of samples with fitting <0.1: 85750
sigmoid_2_param
100%
25384/225384 [03:58<00:00, 943.25it/s]
<function sigmoid_2_param at 0x0000020C7F7C85E0>
R2>0: (225384, 46)
R2> 0.99 6613
Number of samples with fitting <0.1: 88011
sigmoid_3_param
100%
25384/225384 [07:48<00:00, 481.28it/s]
<function sigmoid_3_param at 0x0000020C7F7C8900>
R2>0: (225384, 46)
R2> 0.99 11402
Number of samples with fitting <0.1: 49691
sigmoid_4_param
100%
25384/225384 [28:14<00:00, 133.04it/s]
<function sigmoid_4_param at 0x0000020C7F7C89A0>
R2>0: (225384, 46)
R2> 0.99 15169
Number of samples with fitting <0.1: 123795
 logistic_4_param
25384/225384 [19:55<00:00, 188.60it/s]
```

<function logistic_4_param at 0x0000020C7F7C8B80>

R2>0: (225384, 46) R2> 0.99 16756

Number of samples with fitting <0.1: 68138

114_4_param

100%|

25384/225384 [34:34<00:00, 108.62it/s]

<function II4_4_param at 0x0000020C7F7C8A40>

R2>0: (225384, 46) R2> 0.99 17072

Number of samples with fitting <0.1: 69514

II4R_4_param

100%|

25384/225384 [36:38<00:00, 102.50it/s]

<function II4R_4_param at 0x0000020C7F7C8AE0>

R2>0: (225384, 46) R2> 0.99 16614

Number of samples with fitting <0.1: 75034

logLogist_3_param

100%|

25384/225384 [19:02<00:00, 197.23it/s]

<function logLogist_3_param at 0x0000020C7F7C8C20>

R2>0: (225384, 46) R2>[0]99 16218

Number of samples with fitting <0.1: 90639

CPU times: total: 2h 35min 39s

Wallitime: 2h 35min 29s

df_results.sort_values("R2>0.99")

Out[11]:

	R2>0.9	R2>0.95	R2>0.99
sigmoid_2_param	10372	7236	1409
fsigmoid	10392	7277	1423
sigmoid_3_param	11991	9376	2507
logLogist_3_param	11714	9452	2914
II4R_4_param	10015	8130	3014
sigmoid_4_param	8521	7466	3062
logistic_4_param	10744	8739	3072
II4_4_param	10698	8706	3075

In [12]:

df_results_2

Out[12]:

	R2>0.9	R2>0.95	R2>0.99
fsigmoid	2058	1874	921
sigmoid_2_param	2058	1874	921
sigmoid_3_param	2098	2031	1220
sigmoid_4_param	2045	2017	1527
logistic_4_param	2103	2063	1473
II4_4_param	2103	2063	1473
II4R_4_param	2085	2056	1473
logLogist_3_param	2100	2051	1465

In [13]:

df_results_2.sort_values("R2>0.99")

Out[13]:

	R2>0.9	R2>0.95	R2>0.99
fsigmoid	2058	1874	921
sigmoid_2_param	2058	1874	921
sigmoid_3_param	2098	2031	1220
logLogist_3_param	2100	2051	1465
logistic_4_param	2103	2063	1473
II4_4_param	2103	2063	1473
II4R_4_param	2085	2056	1473
sigmoid_4_param	2045	2017	1527

In [14]:

df_results_3

Out[14]:

	R2>0.9	R2>0.95	R2>0.99
fsigmoid	53091	32241	6638
sigmoid_2_param	53063	32194	6613
sigmoid_3_param	70557	46861	11402
sigmoid_4_param	53846	42088	15169
logistic_4_param	73125	52855	16756
ll4_4_param	73134	53103	17072
II4R_4_param	67848	49454	16614
logLogist_3_param	72883	53643	16218

In [15]:

df_results_3.sort_values("R2>0.99")

Out[15]:

	R2>0.9	R2>0.95	R2>0.99
sigmoid_2_param	53063	32194	6613
fsigmoid	53091	32241	6638
sigmoid_3_param	70557	46861	11402
sigmoid_4_param	53846	42088	15169
logLogist_3_param	72883	53643	16218
II4R_4_param	67848	49454	16614
logistic_4_param	73125	52855	16756
II4_4_param	73134	53103	17072

In []: