

Session: minf_201200059_sm_dataset_198cpds, Model: P-gp_pEC50

Mon October 15 2012, 20:45

Data set: minf_201200059_sm_dataset_198cpds

Modeled property: pEC50

Modeling technique: Gaussian Processes

Model statistics:

	Number	Rsqr	RMSE
TRN	140	0.9258	0.2977
VAL	29	0.7036	0.614
TEST	29	0.6033	0.5019

Parameters used:

Set split:

- Training set size: 70%
- Validation set size: 15%
- Clustering with tanimoto level: 0.7

Descriptor pre-selection:

- Threshold for minimum occurrence: 4%
- Threshold for minimum standard deviation: 0.0005
- Threshold for maximum correlation between descriptors: 0.95

Descriptors remaining after pre-selection: 132

Descriptors used in the model: 30

Model details:

Theta1: 41.22486877

Theta2: 24.94758797

Theta3: 0.6999999881

Descriptor	Length scale
logP	5.551753
Vx	3073.761
ERTLNotPSA	823.5327
ERTLNoSPtPSA	941.5363
HBD-prof	28.28905
CH1Aa	3.436333
NRB	133.6726
aromCl	14.76831
lipovolume	50.8041
phenolic-tautomer	4.078263
ringat	325.6476
nC(sp2)	260.3202
nC(sp3)	21.12252
nCO	51.02331
nNprot	2.23126
sssCH	22.98964
aasC	28.91639
sF	40.01468
UB	306.0531
f004	51.90596
f440	48.99949
q040	420.3657
q137	294.545
q139	225.9982
q155	211.5906

q192
q300
q453
q458
q481

271.8256
77.18587
236.655
259.4429
721.1013