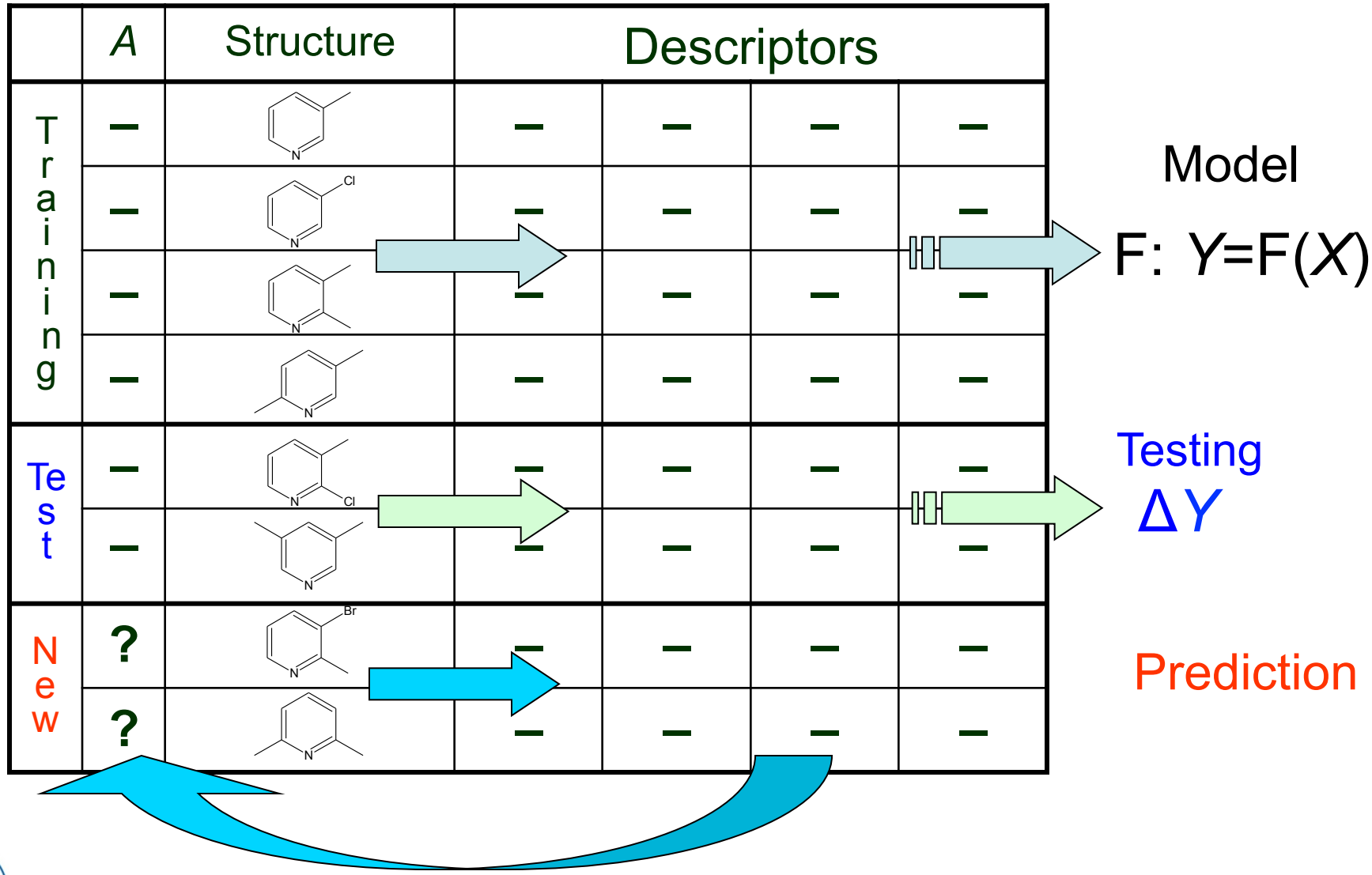
Machine-Learning Methods in Property Predictions: *Quo Vadis*?

Igor I. Baskin

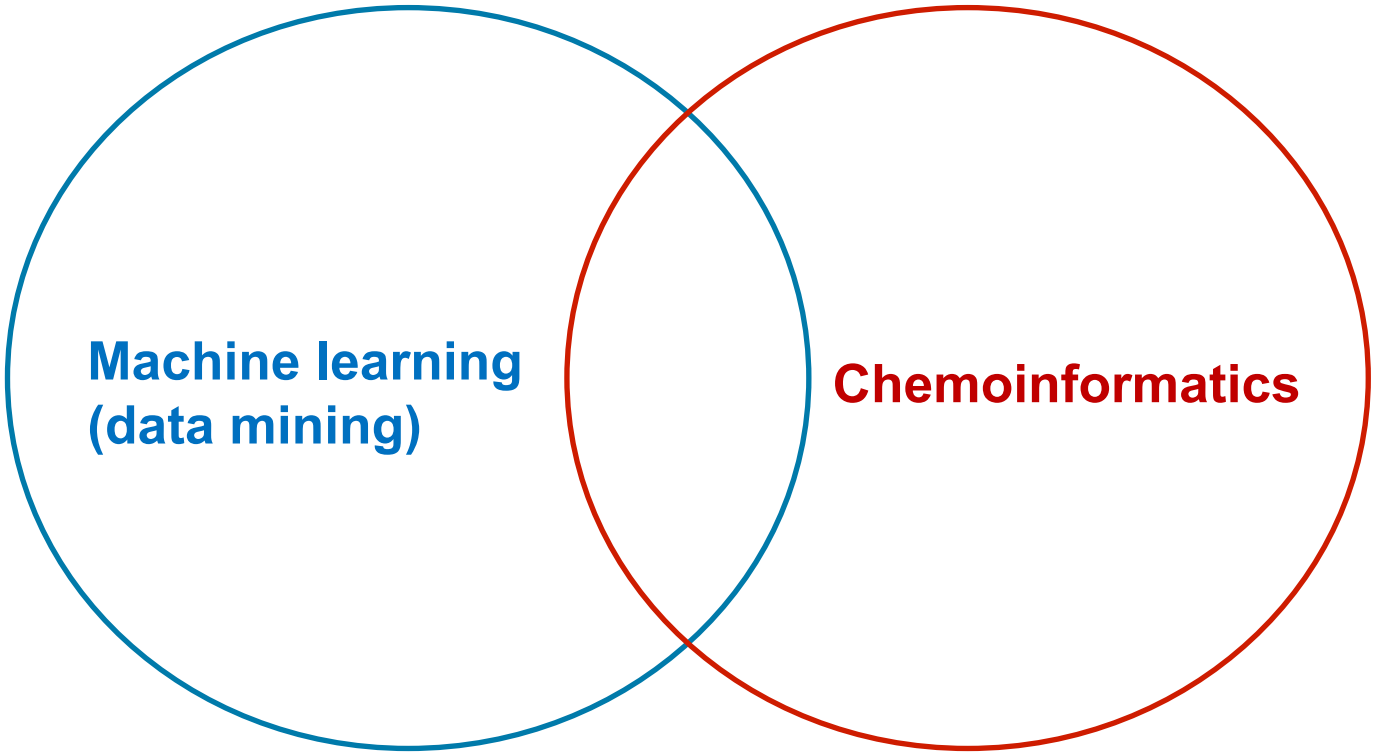
Lomonosov Moscow State University
RUSSIA

Физический факультет
Московский государственный университет имени М.В. Ломоносова

General Workflow for QSAR Modiling in Chemoinformatics



Machine Learning and *Chemoinformatics*: different but overlapping fields



Chemometrics

- Chemometrics is what chemometricians do.
- Chemometricians are people who drink beer and steal ideas from statisticians

Svante Wold

Chemoinformatics
~~Chemometrics~~

Chemoinformatics chemoinformaticians

- C~~hemometrics~~ is what ~~chemometricians~~ do

Chemoinformaticians

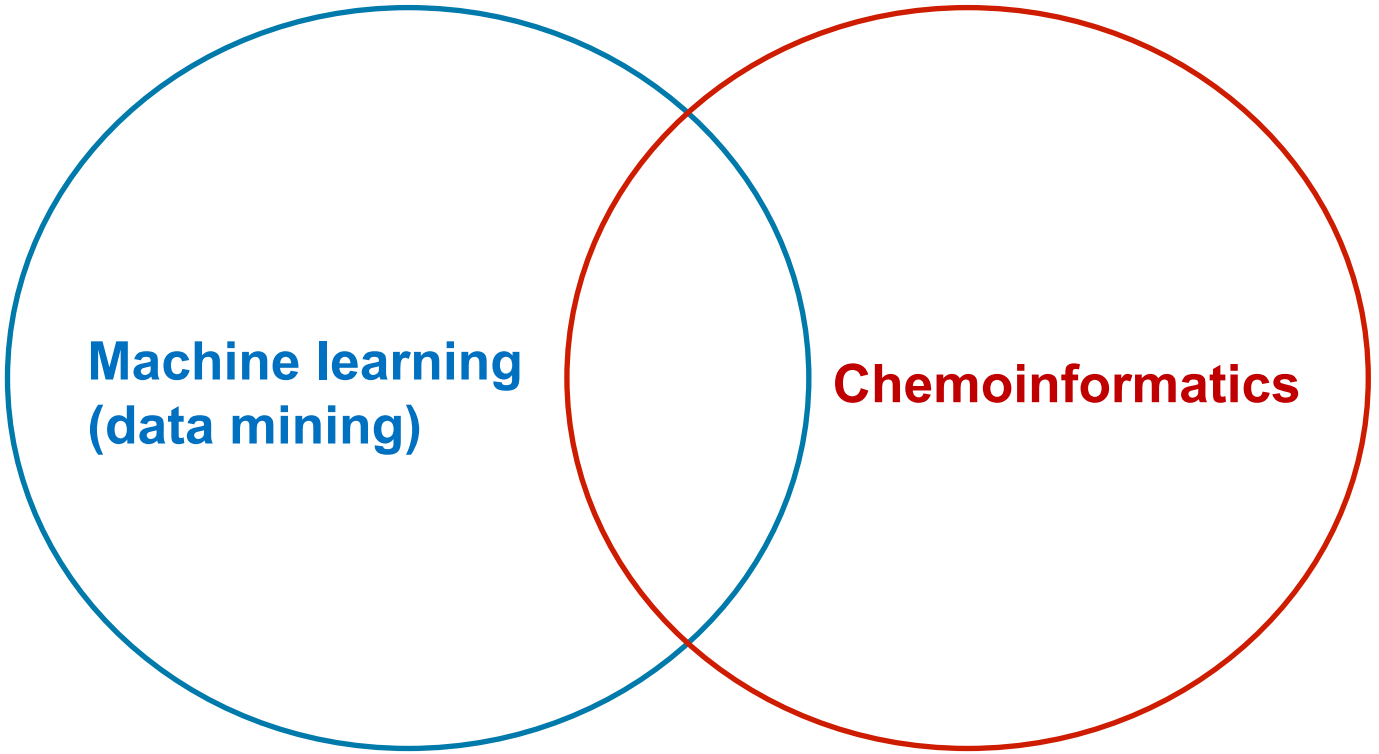
- C~~hemometricians~~ are people who drink beer (??)

borrrow machine-learners

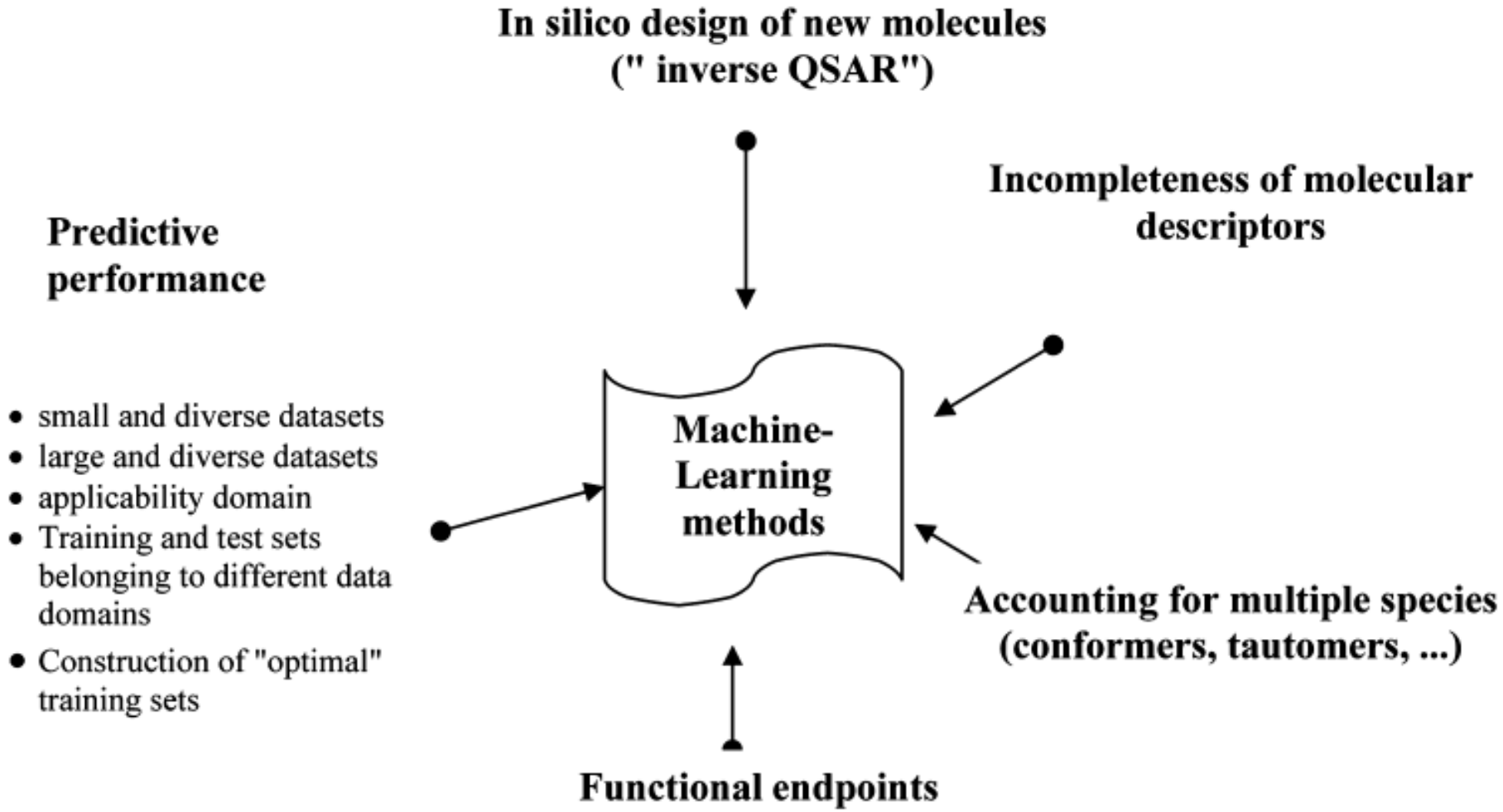
and ~~steal~~ ideas from ~~statisticians~~

.

Machine Learning and *Chemoinformatics*: different but overlapping fields

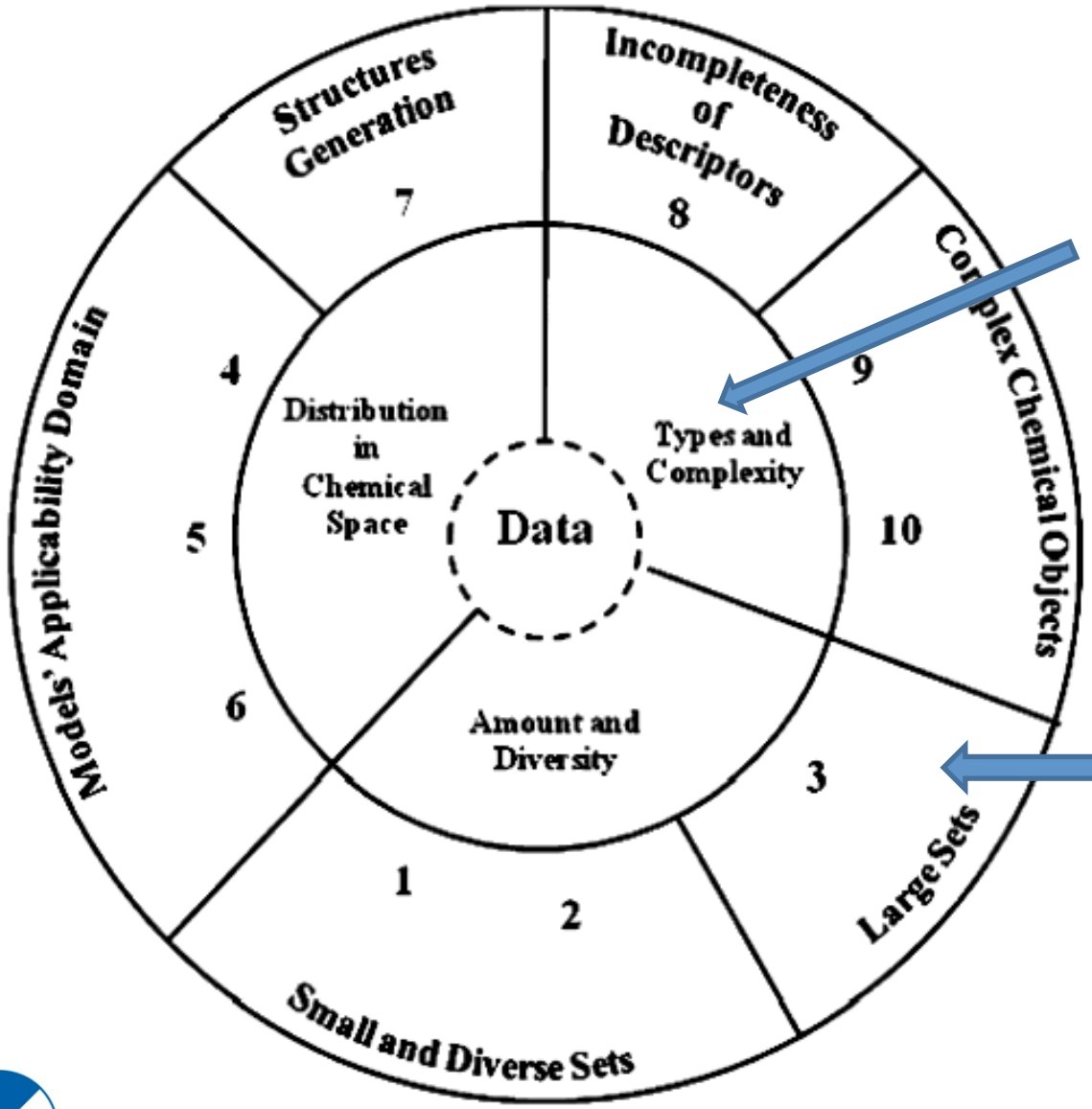


Main Challenges of Machine-Learning Methods in Chemoinformatics



A.Varnek, I. Baskin. *J. Chem. Inf. Mod.* **2012**, 52 (6), 1413-1437

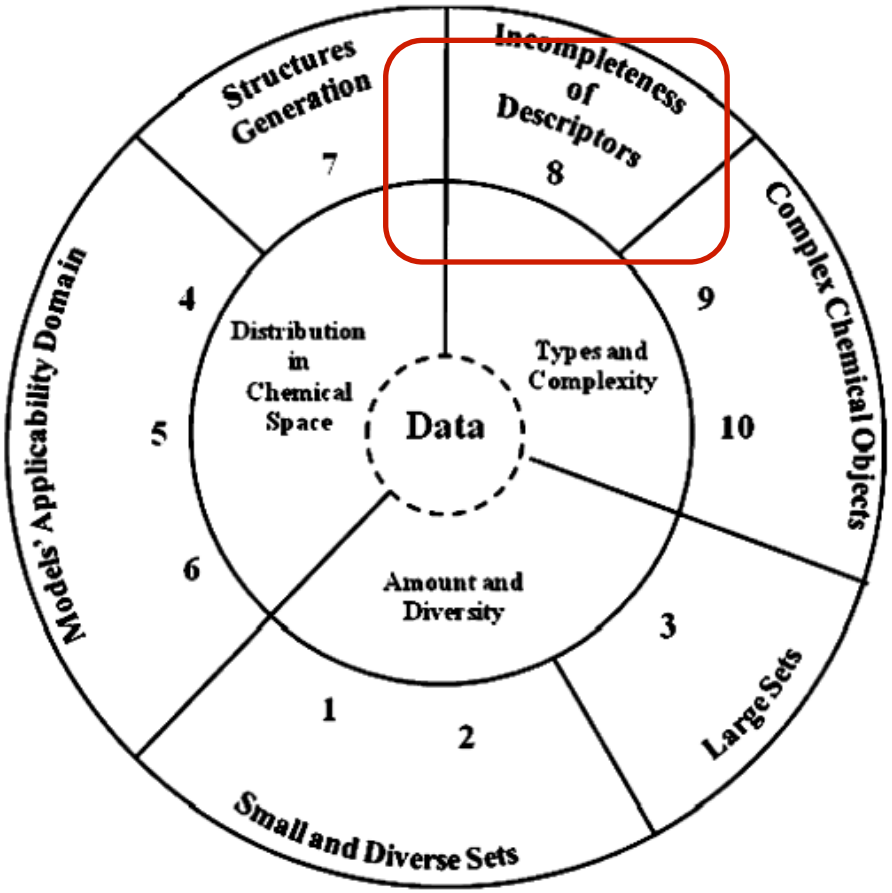
Guide to Choose Machine Learning Method to solve Chemical Problems



Different features of the data (*inner circle*)

Challenges of chemoinformatics (*outer circle*)

Machine Learning on Molecular Graphs



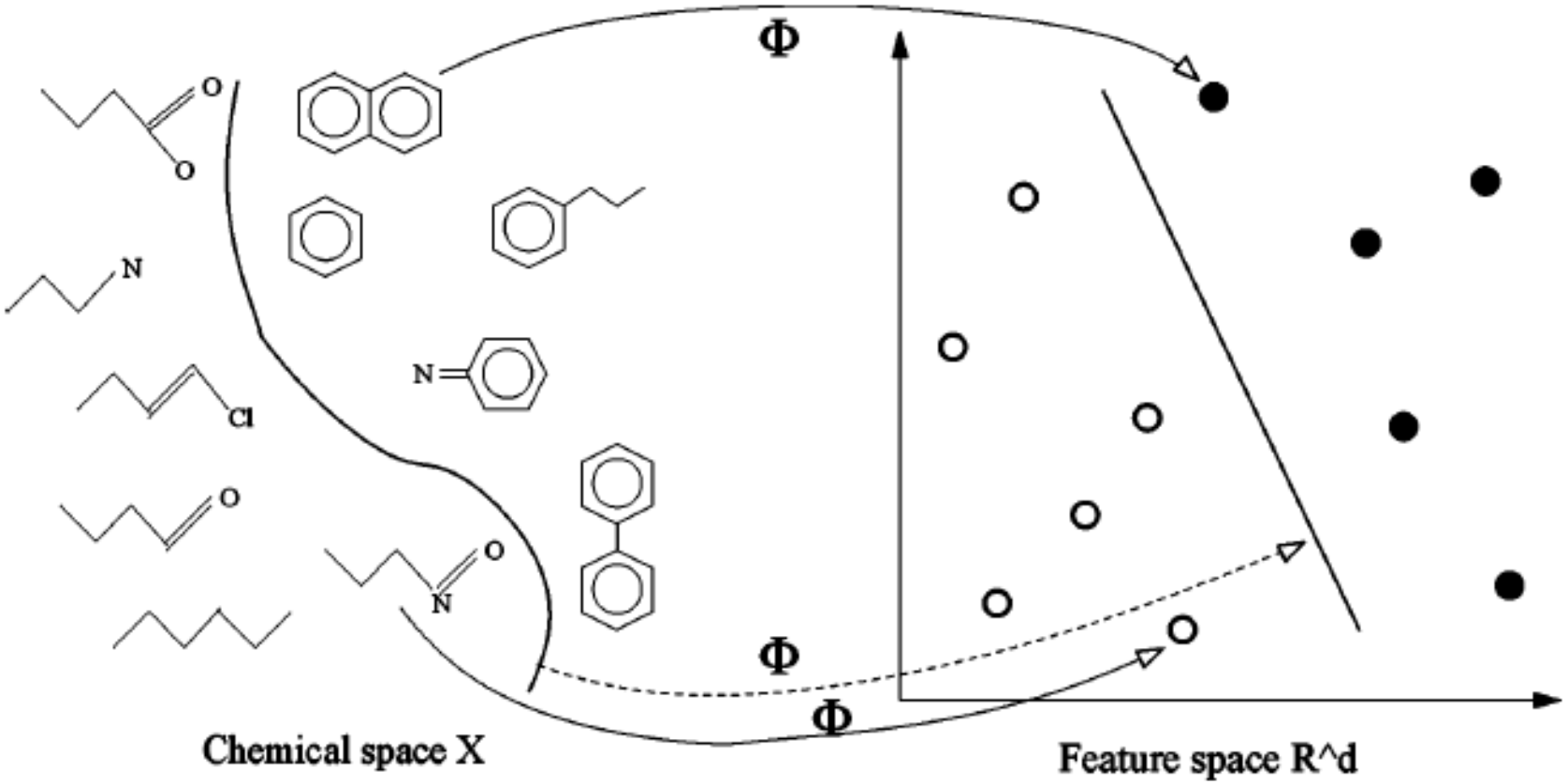
Is it possible to build a model directly on molecular graphs instead of using fixed-sized vectors of descriptors?

Graph → **Model** → **Property**

- Graph mining with special architectures of neural networks
- (Sub)Graph mining
- Graph kernels
- Inductive learning programming
- Symmetry-invariant machine learning with local features
- Energy-based learning
- etc

- G.Bakir, T.Hofmann, B.Schoelkopf, A.J.Smola, B.Taskar, S.V.N.Vishwanathan. Predicting Structured Data; The MIT Press:Cambridge, MA, 2007.
- D.J.Cook, L.B.Holder. Mining Graph Data; Wiley-Interscience: Hoboken, NJ, 2007.

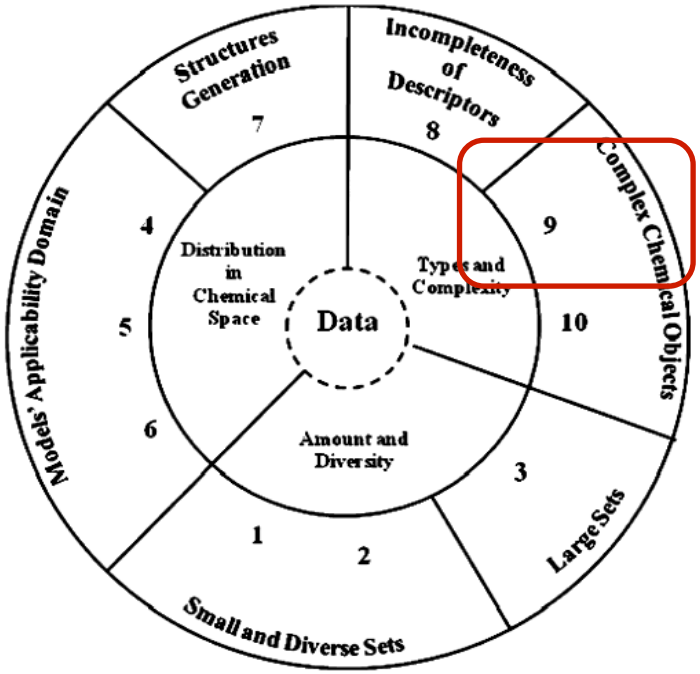
Machine Learning on Graph Kernels



$$< \Phi(x), \Phi(x') > = K(x, x')$$

- M.Rupp, G.Schneider. *Mol. Inf.* **2010**, *29* (*4*), 266−273

Multi-Instance Learning

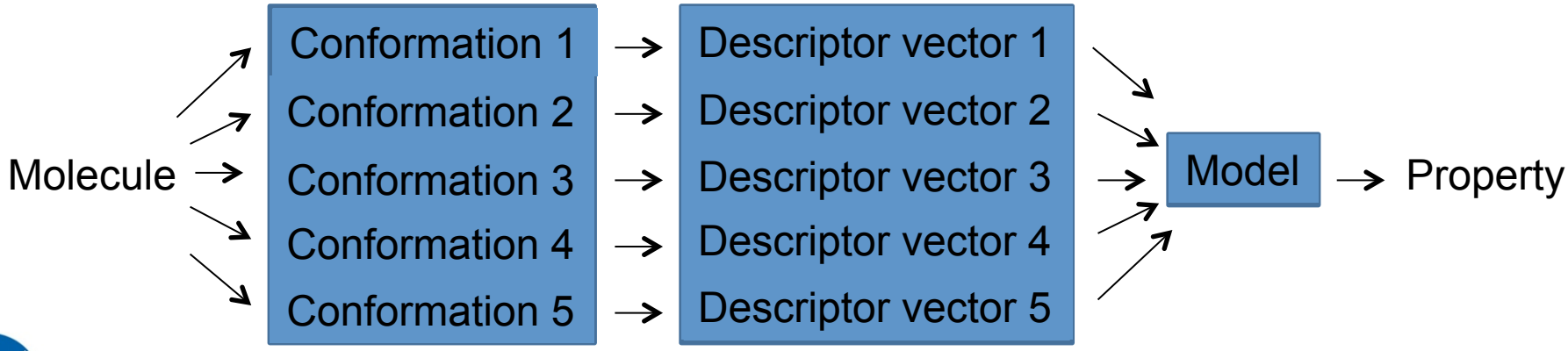


Representing molecule as a number of conformers, tautomers and ionization forms, …

Every object represents an ensemble (so-called bag) of instances, each of which is described by a fixed-sized vector of descriptors.

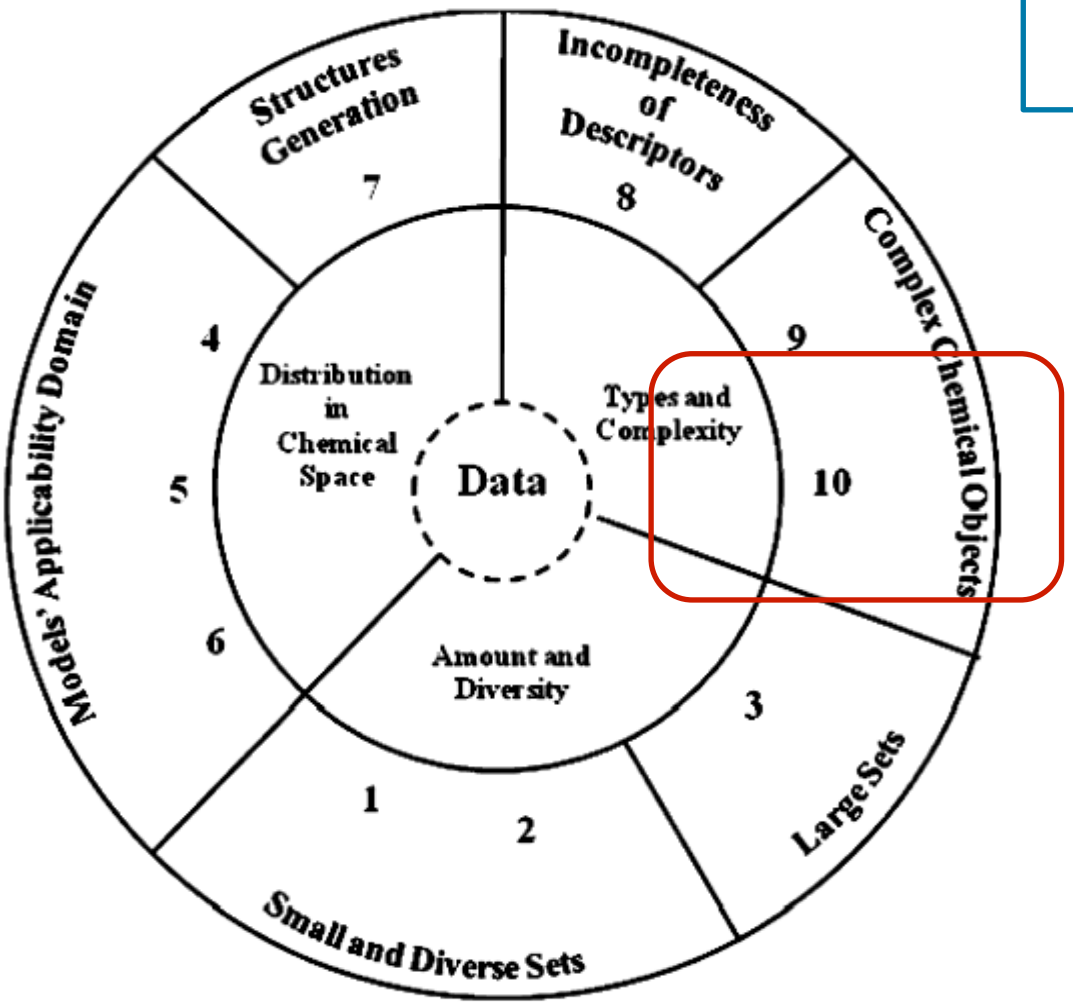
Instances (conformations, tautomers, etc)

Bag of feature vectors (descriptor vectors)

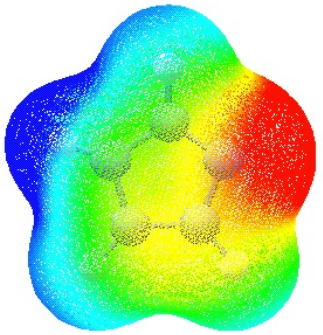


T.G.Dietterich, R.H.Lathrop, T. Lozano-Pérez. *Artif. Intell*. **1997**, *89* (*1−2*), 31−71

Functional Data Analysis



FDA allows one to build models for molecules represented by functions?



Objects represented by functions

↓

Models

↓

Properties

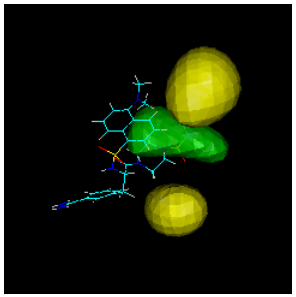
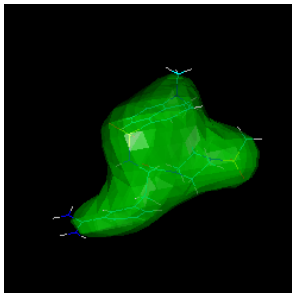
Ramsay, J. O.; Silverman, B. W. *Functional Data Analysis*. 2nd ed.; Springer: NY, USA, 2005

Continuous Molecular Fields (CMF)

Continuous Molecular Fields approach describes molecules by ensemble of continuous functions (***molecular fields***), instead of finite sets of molecular descriptors**.** CMF is kernel-based method.

traditional QSAR $Activity = F(X) = \sum c_i x_i$

CMF $Activity = F[\mathrm{X}(\mathbf{r})] = \int C(\mathbf{r})\mathrm{X}(\mathbf{r})d\mathbf{r}$

$Activity = \int$  $\cdot$  $d\mathbf{r}$

C(r) **X(r)**

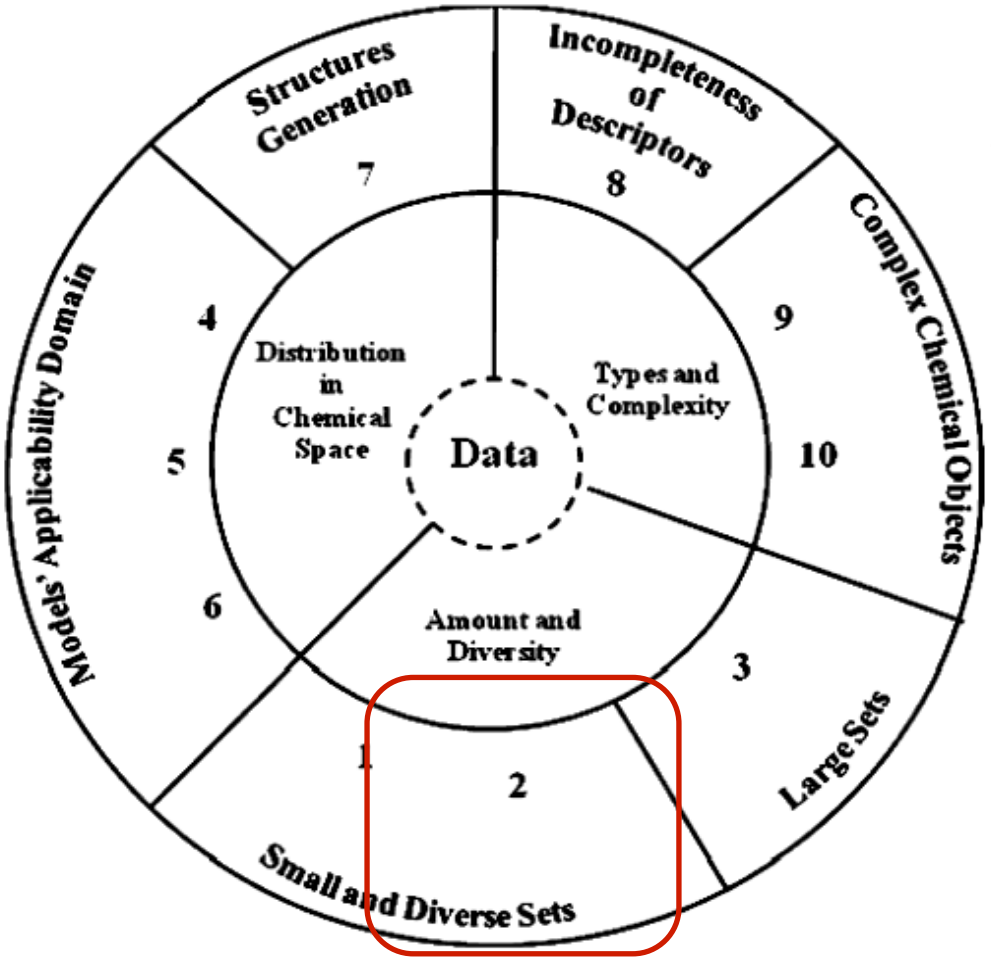
Calculated using special kernels for molecular fields

Gaussian functions approximation of molecular fields

http://sites.google.com/site/conmolfields/

I.I.Baskin, N.I. Zhokhova. *J. Comput.-Aided Mol. Des.* **2013**, *27* (*5*), 427-442

Inductive Knowledge Transfer

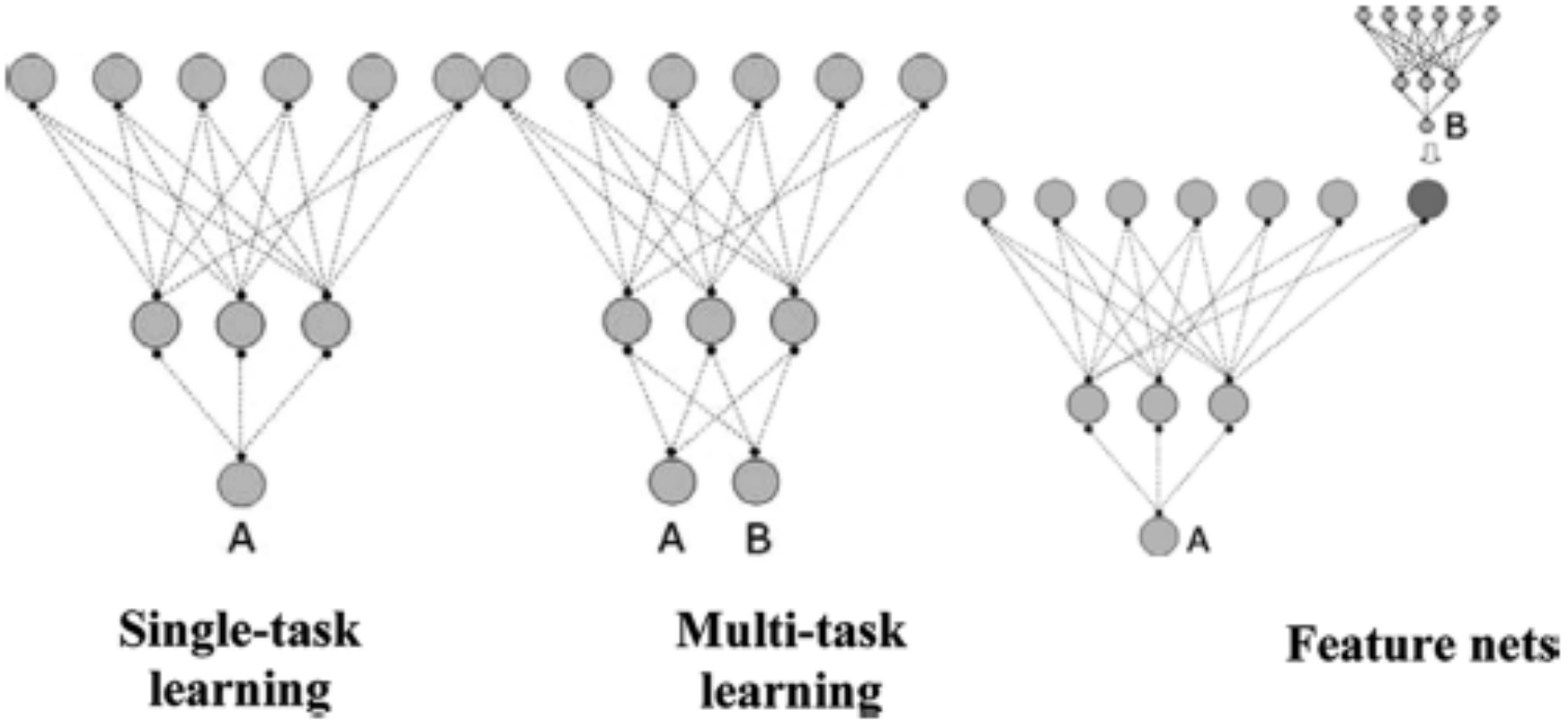


(inductive bias, lifelong learning, learning to learn, collaborative filtering, multi-task learning etc)

Transfer of information from one model, usually trained on sufficiently large dataset, to another model trained on small dataset

•*Learning to Learn; S.Thrun,* L.Y.Pratt, Eds.; Kluwer Academic Publishers: Boston, MA, 1998

Interference of Models (Inductive Knowledge Transfer)

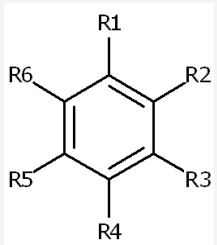


A.Varnek, C.Gaudin, G.Marcou, I.Baskin, A.K.Pandey, I.V.Tetko. *J. Chem. Inf. Mod.* **2009**, *49* (*1*), 133-144.

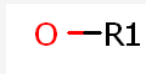
Partition coefficients air-tissue

The blood:air partition coefficient (PC) is an important determinant of the distribution of volatile organic chemicals (VOCs).

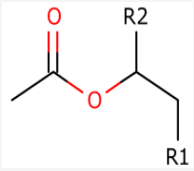
	Tissue	N
Human	blood	139
	fat	42
	brain	36
	liver	34
	muscle	39
	kidney	34
Rat	fat	99
	brain	59
	liver	100
	muscle	97
	kidney	27



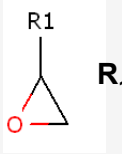
R_1=Me,Et,Pr,iPr,
CH_2=CH_2CH_3,CH_2=CH_2,F,Cl,Br
R_2,R_3=H,Me,F
R_4=H,Me,CH_2=CH_2,F,CF_3
R_5=H,CH_2=CH_2,CH_3,F
R_6=H,CH_3,F,Cl



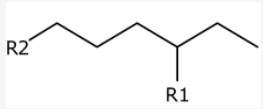
R_1=Me, Et, Pr, iPr, Bu, iBu, C_5H_{11},tBu



R_1=Me,Et,Pr, iBu, iPr
R_2=Me



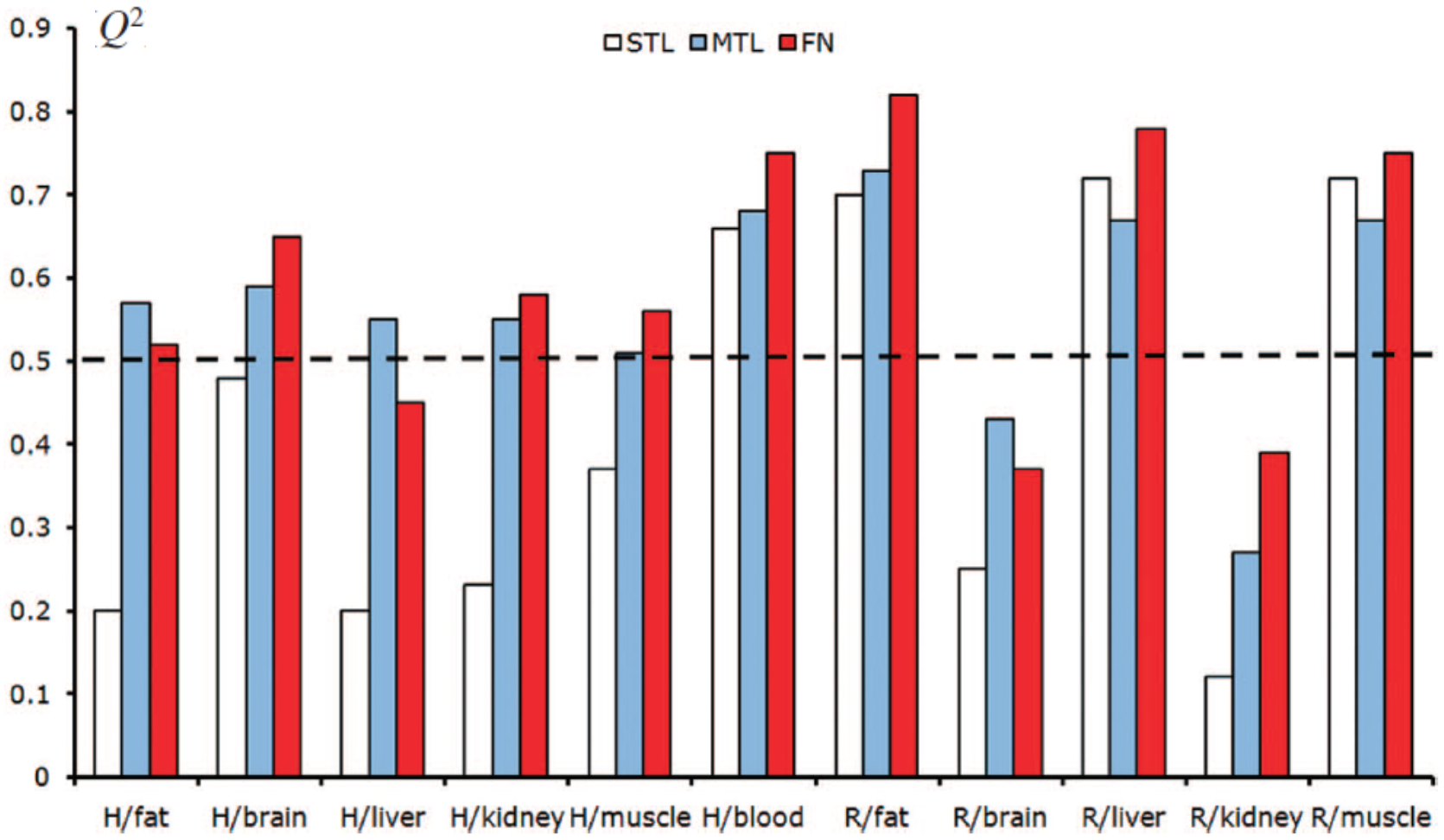
R_1=H,CN,CH=CH_2



R_1=H,Me,OH
R_2=Me,Pr,Bu,OH,SH

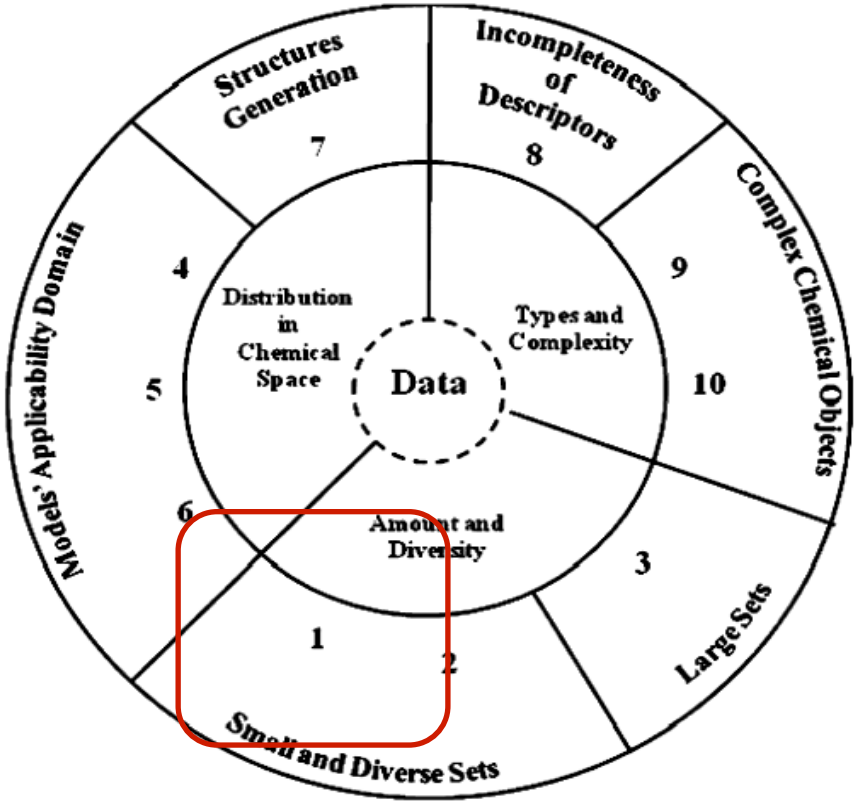
A. Katritzky, A. Varnek et al. *Bioorganic & Medicinal Chemistry*, **2005,** *13*,6450–6463

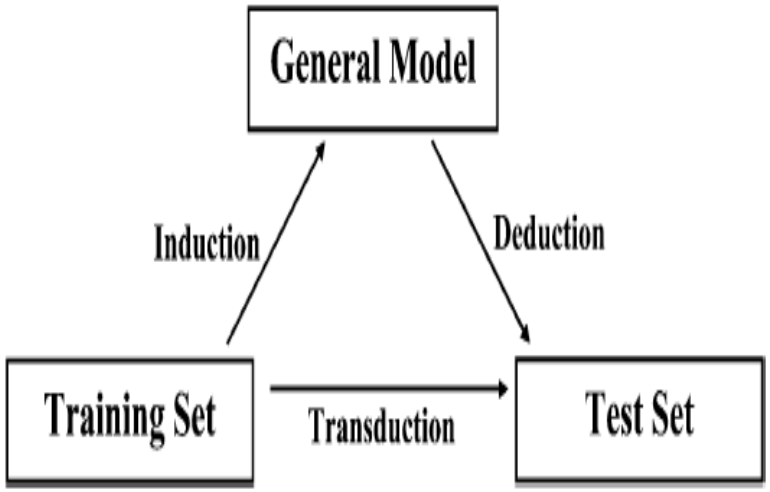
Inductive Knowledge Transfer (Modeling Tissue-Air Partition Coefficients)



A.Varnek, C.Gaudin, G.Marcou, I.Baskin, A.K.Pandey, I.V.Tetko. *J. Chem. Inf. Mod.* **2009**, *49* (*1*), 133-144.

Transductive (Semi-Supervised) Machine Learning

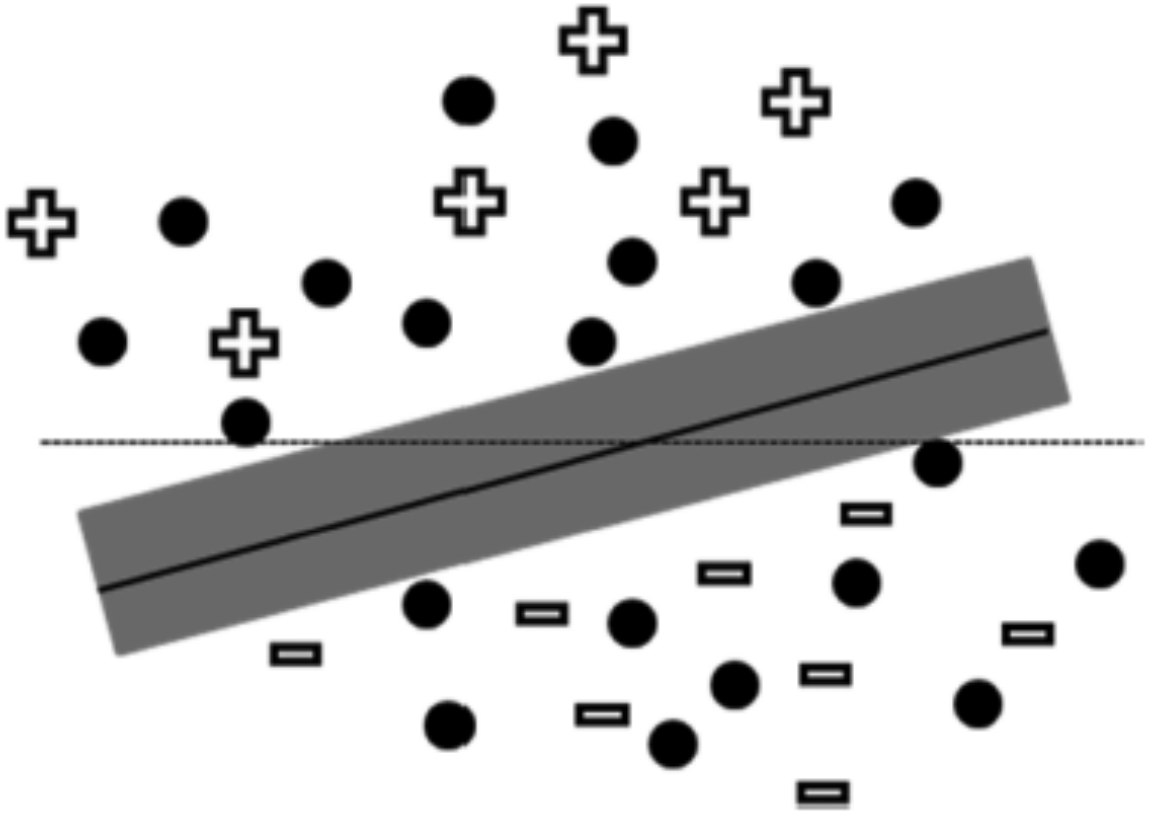




Transductive modeling is used to build the models specifically oriented toward the best prediction performance on a particular test set instead of developing general models to be applied to any test set

V. Vapnik, *Statistical Learning Theory*, Wiley-Interscience, New York, **1998**.

Object Separation in SVM and TSVM

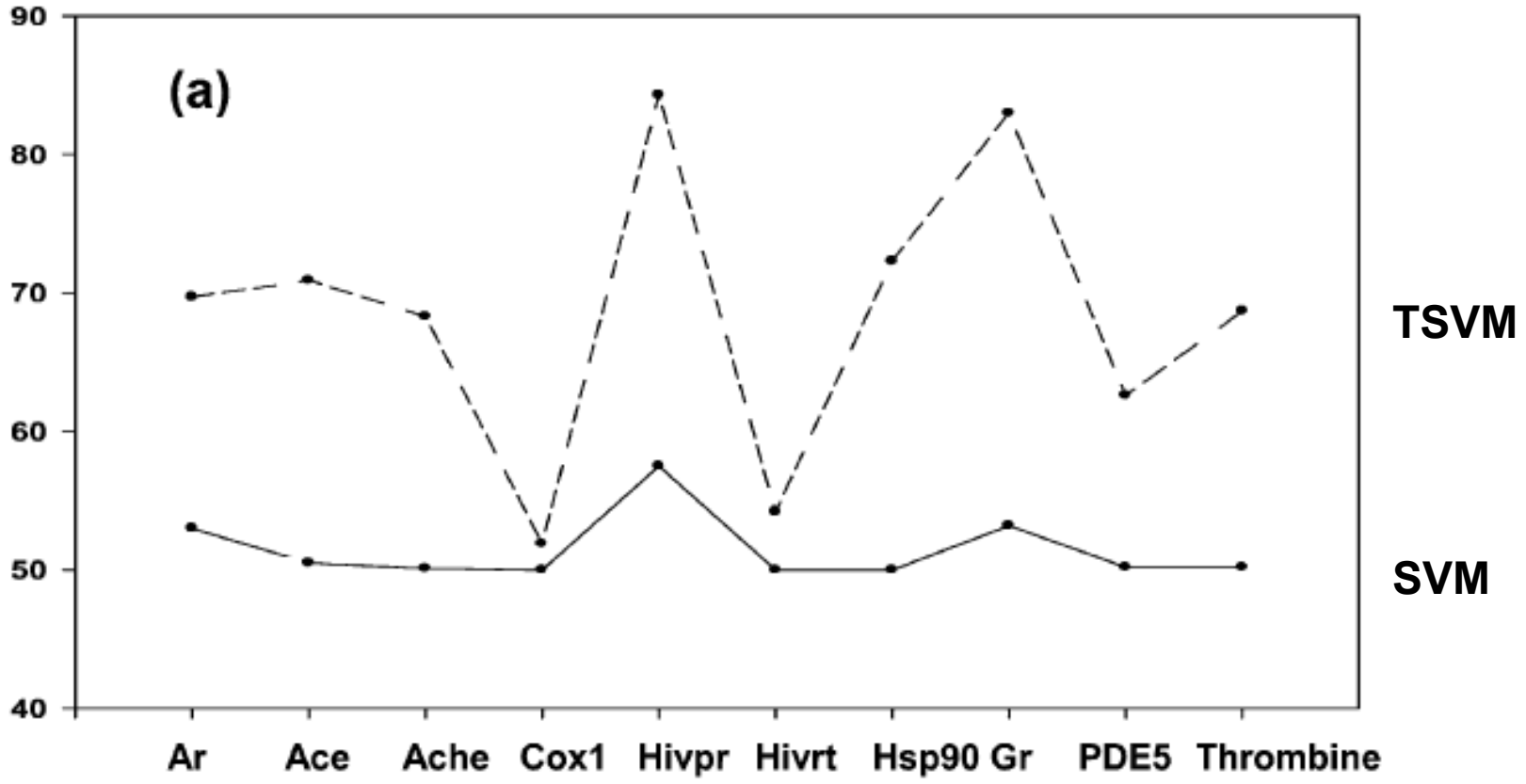


Labeled training set examples are depicted as signs - and +,. Unlabeled test set examples are shown as bold dots.

T. Joachims, in International Conference on Machine Learning (ICML) (Ed: M. Kaufmann), Bled, Slovenia, 1999, pp. 200–209.

Prediction Performance (Balanced Accuracy) of SVM vs TSVM Models

(Training sets consist of 5 active and 50 inactive compounds)

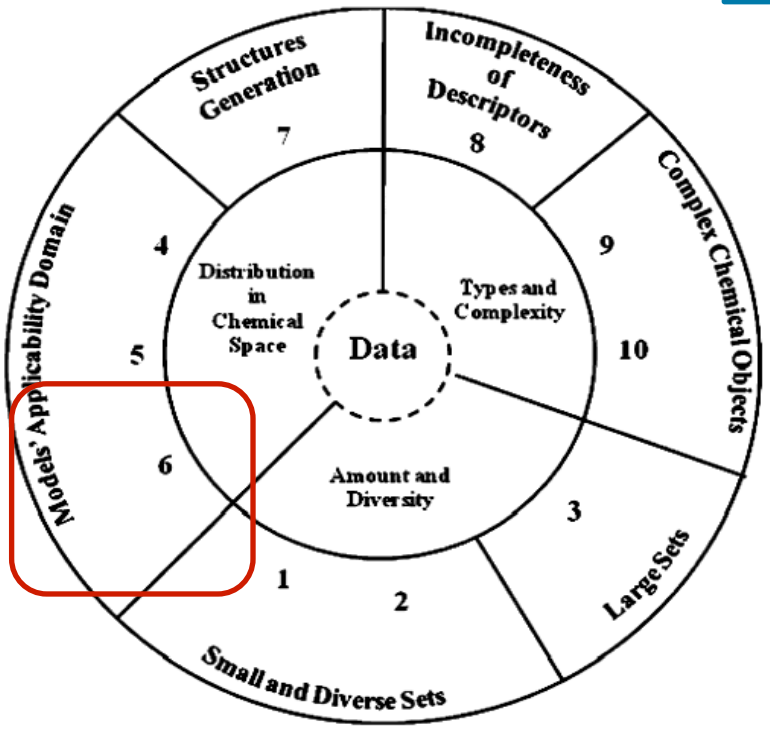


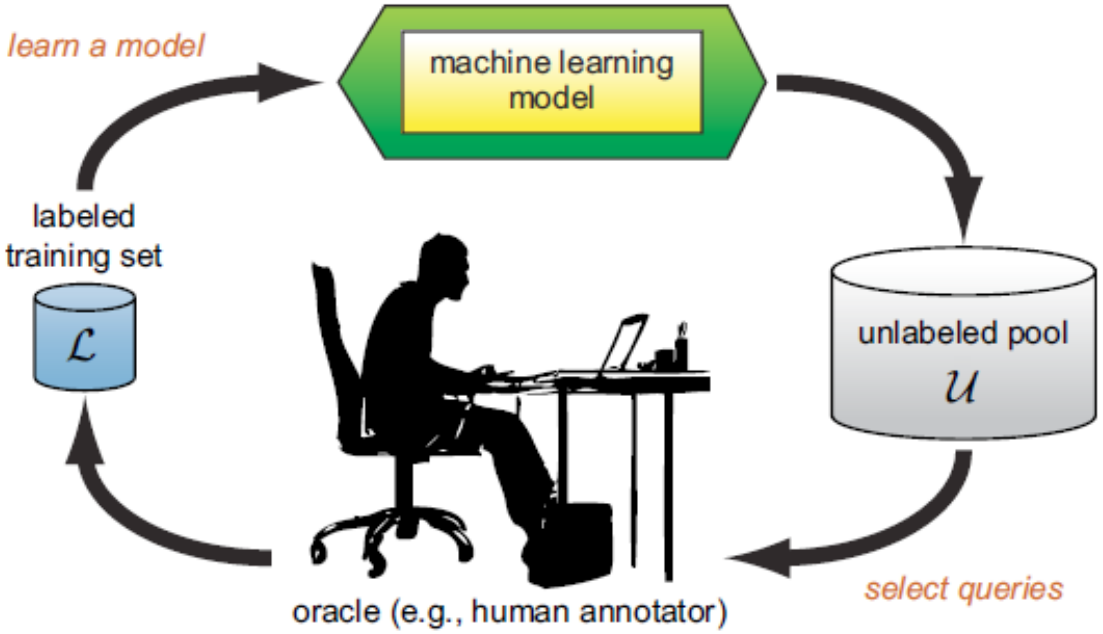
Transductive effect is the difference in prediction performance between transductive and inductive models

E.Kondratovich, I.I.Baskin, A.Varnek. *Mol. Inf.* **2013**, *32* (*3*), 261-266

Active Learning

Active learning helps to form “optimal” training sets

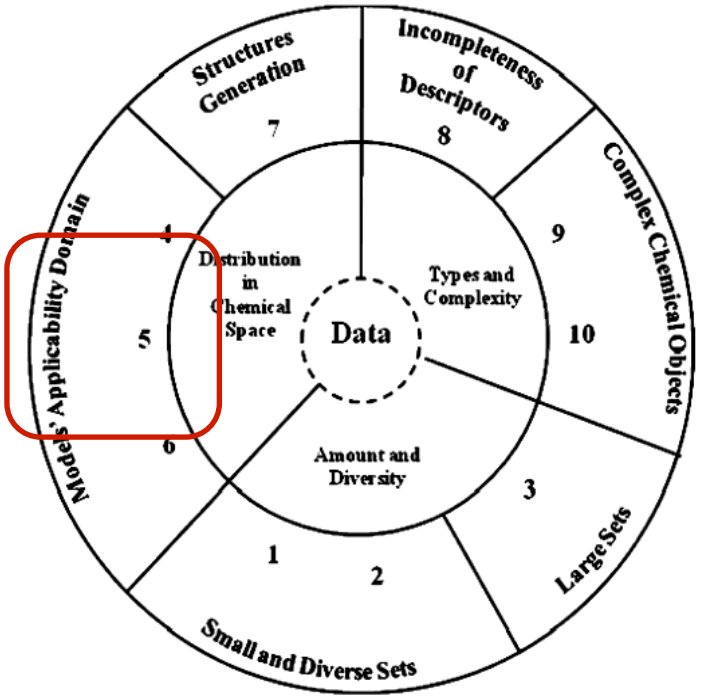




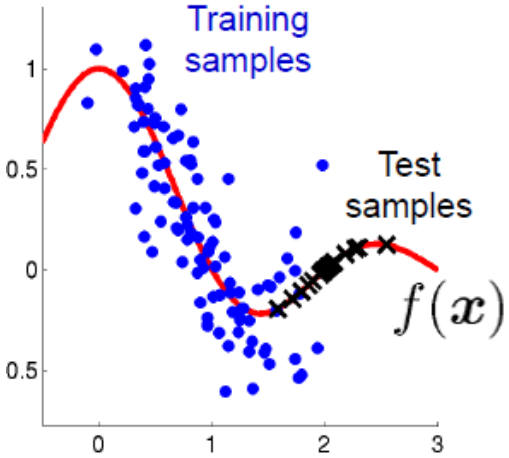
In each learning iteration, the most “useful” compound is selected from a pool, studied in experiment and added to the training set followed by model rebuilding

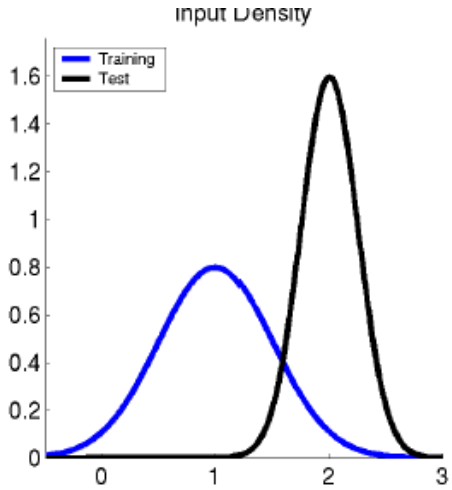
- Burr Settles. Active Learning Literature Survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison. 2009 (http://active-learning.net)
- Y.Fujiwara, Y.Yamashita, T.Osada et al. *J. Chem. Inf. Model.* **2008**, *48* (*4*), 930−940

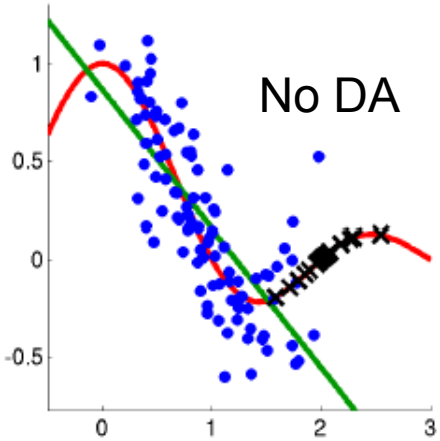
Domain Adaptation

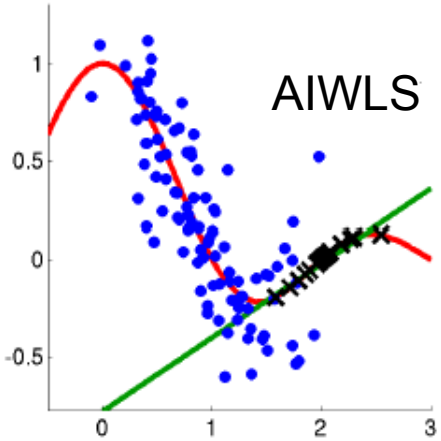


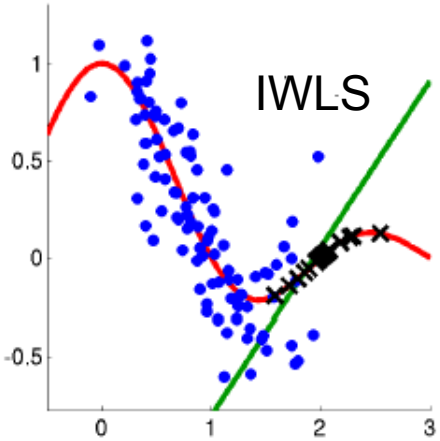
What to do if the training and the test sets are drawn from different distributions?





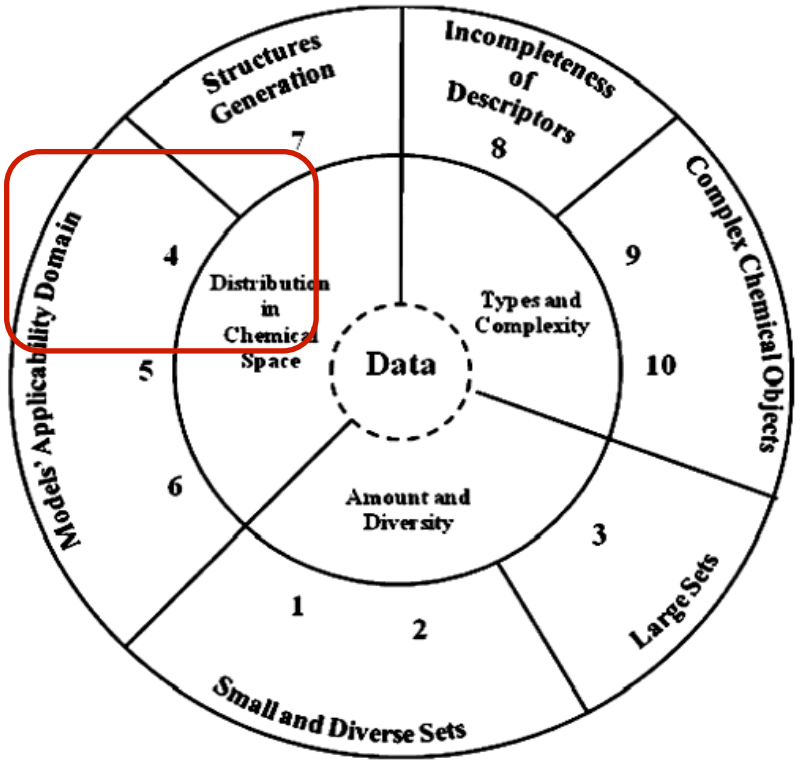




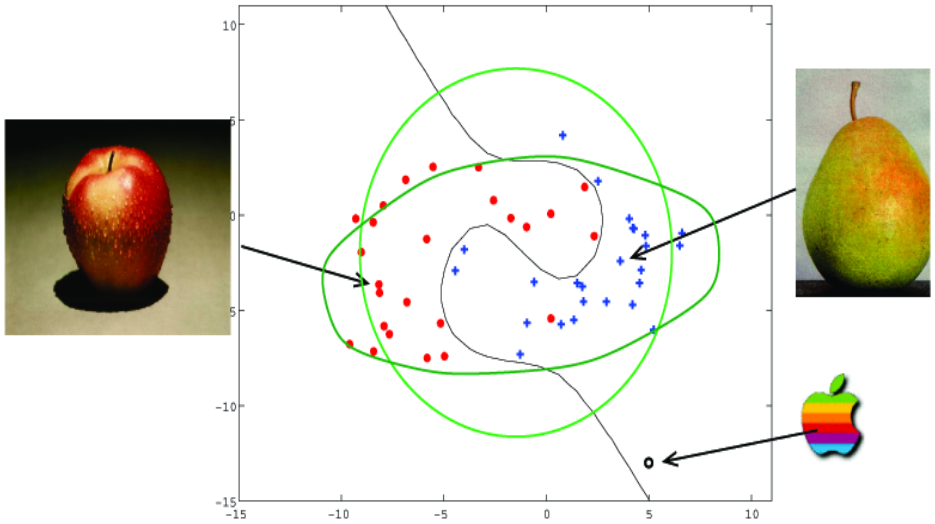


M.Sugiyama, M.Krauledat, K.-R.Mueller. *J. Mach. Learn. Res.* **2007**, *8*, 985−1005.

One-Class Classification (Novelty Detection)



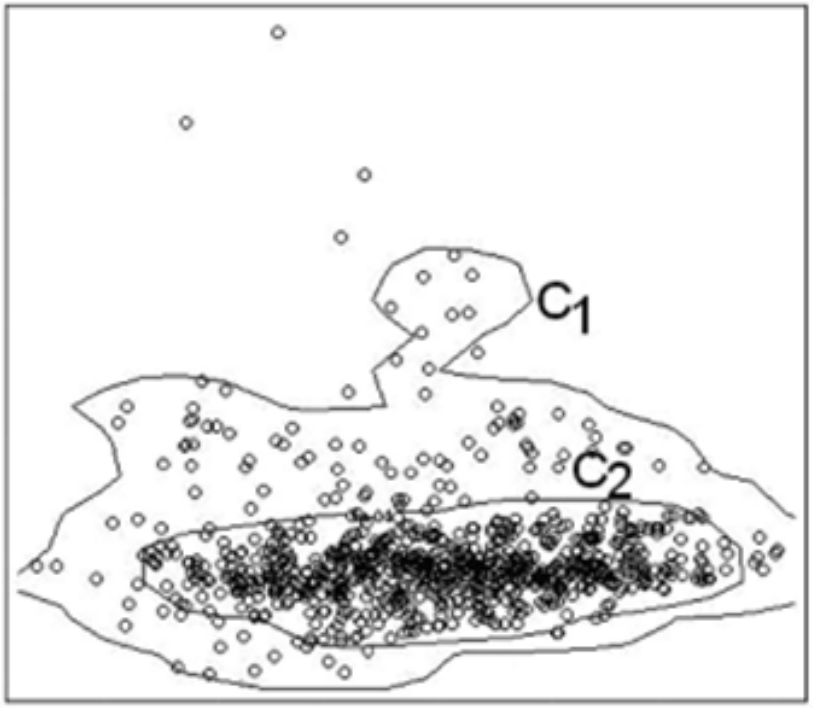
How to build classification models without counterexamples?

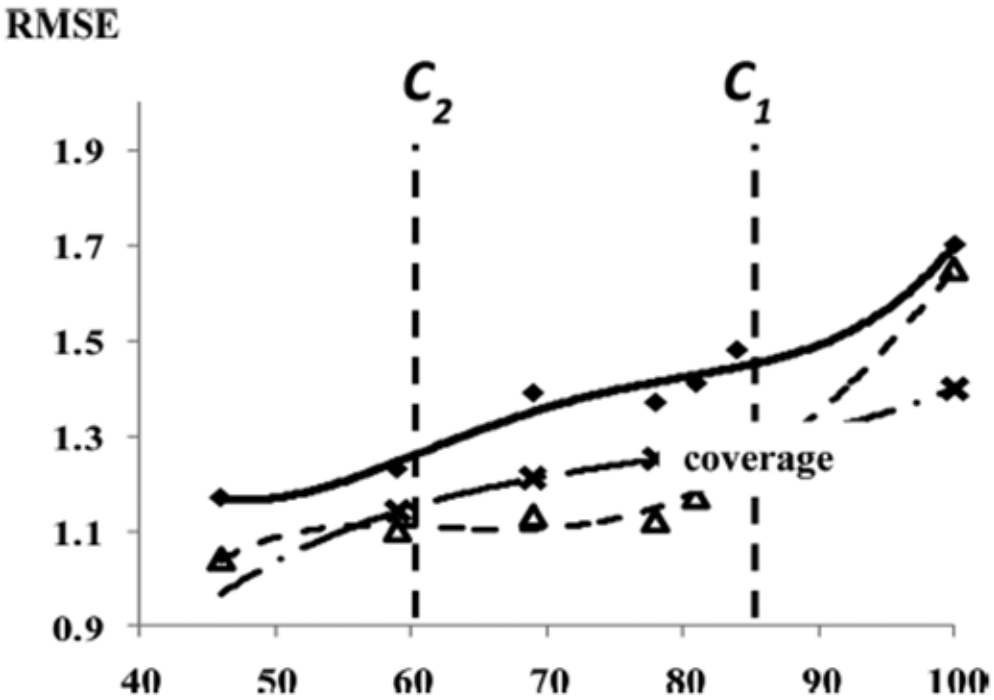


One-class classification (or novelty detection) methods allows one to build classification models without counterexamples. In contrast to conventional (two-class) classification, one-class classification tends to describe one single class of objects (*target class objects*), and distinguish it from all other objects (*outliers*).

D.M.J. Tax, Doctor Thesis, Technische Universiteit Delft, The Netherlands, **2001**

One-Class Classification (OCC) Approach to Defining Model Applicability Domain (AD)

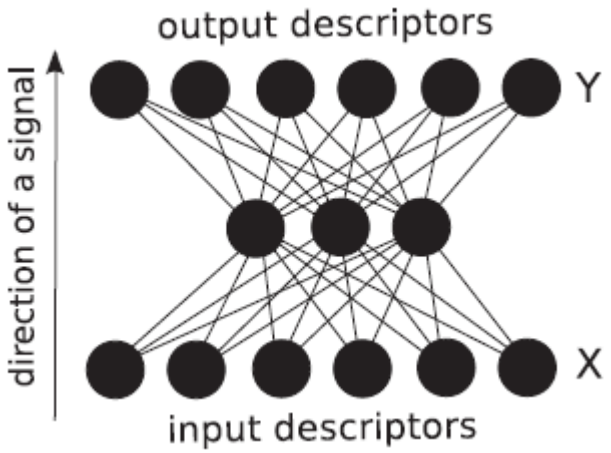




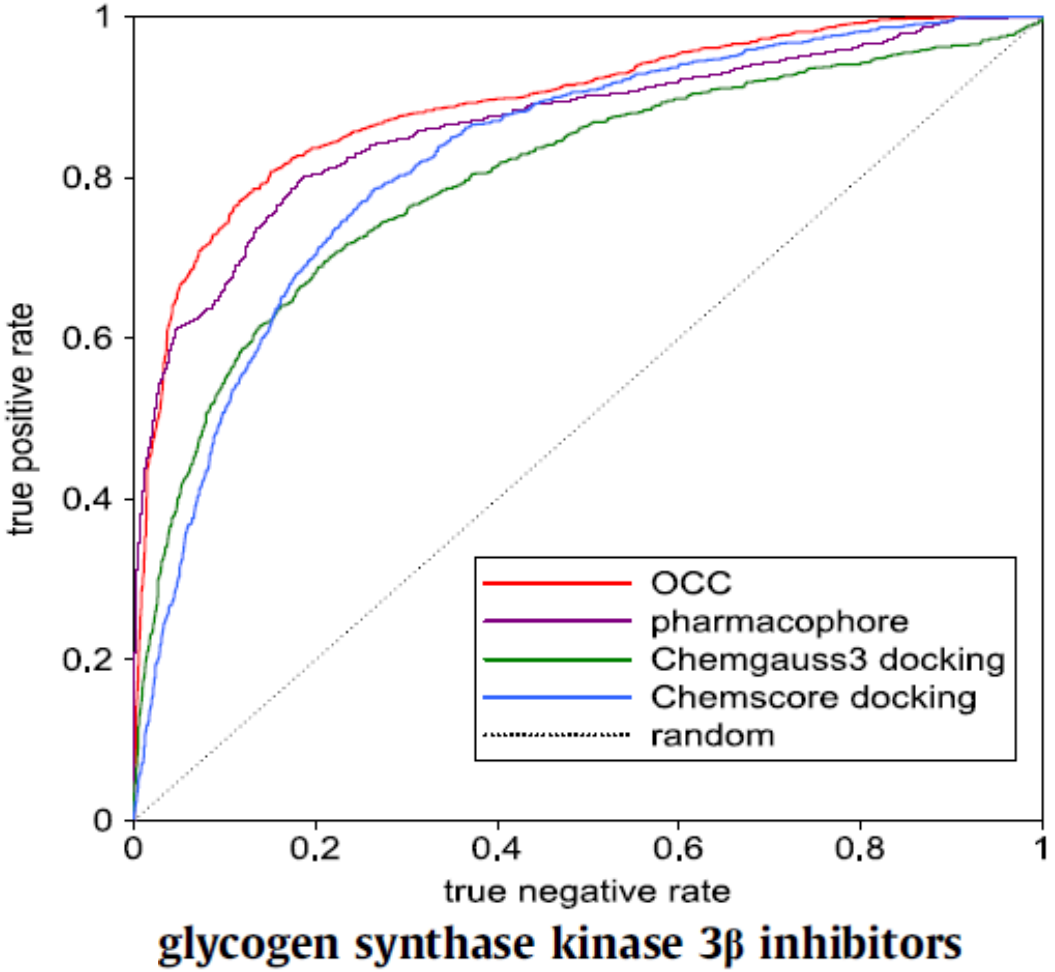
QSPR modeling of stability constants for of Ca^{2+} , Sr^{2+} and Ba^{2+} with organic ligands

I.I.Baskin, N.Kireeva, A.Varnek. *Mol. Inf.* **2010**, *29* (*8-9*), 581-587.

Virtual Screening Based on One-Class Classification Using Auto-Encoder Neural Network

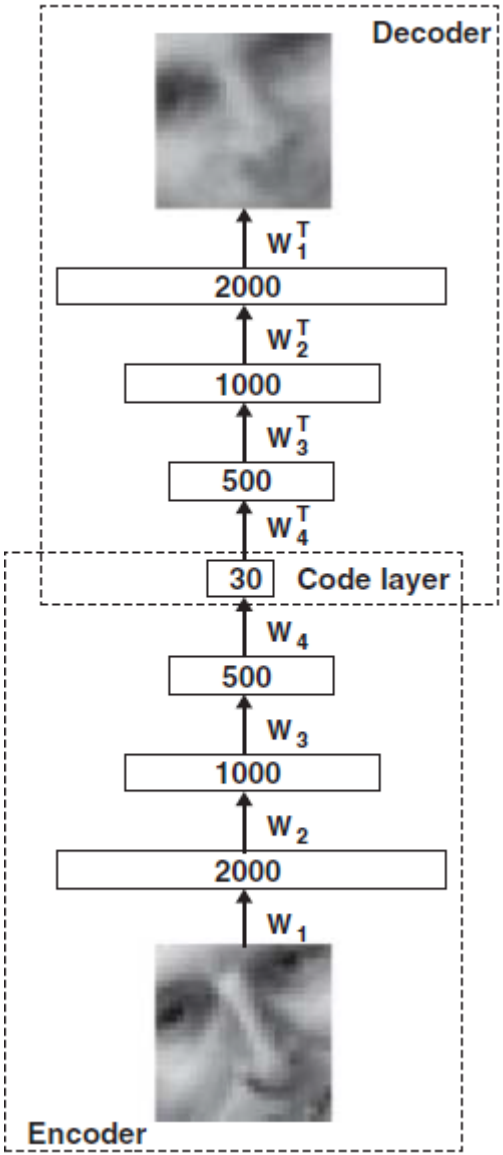


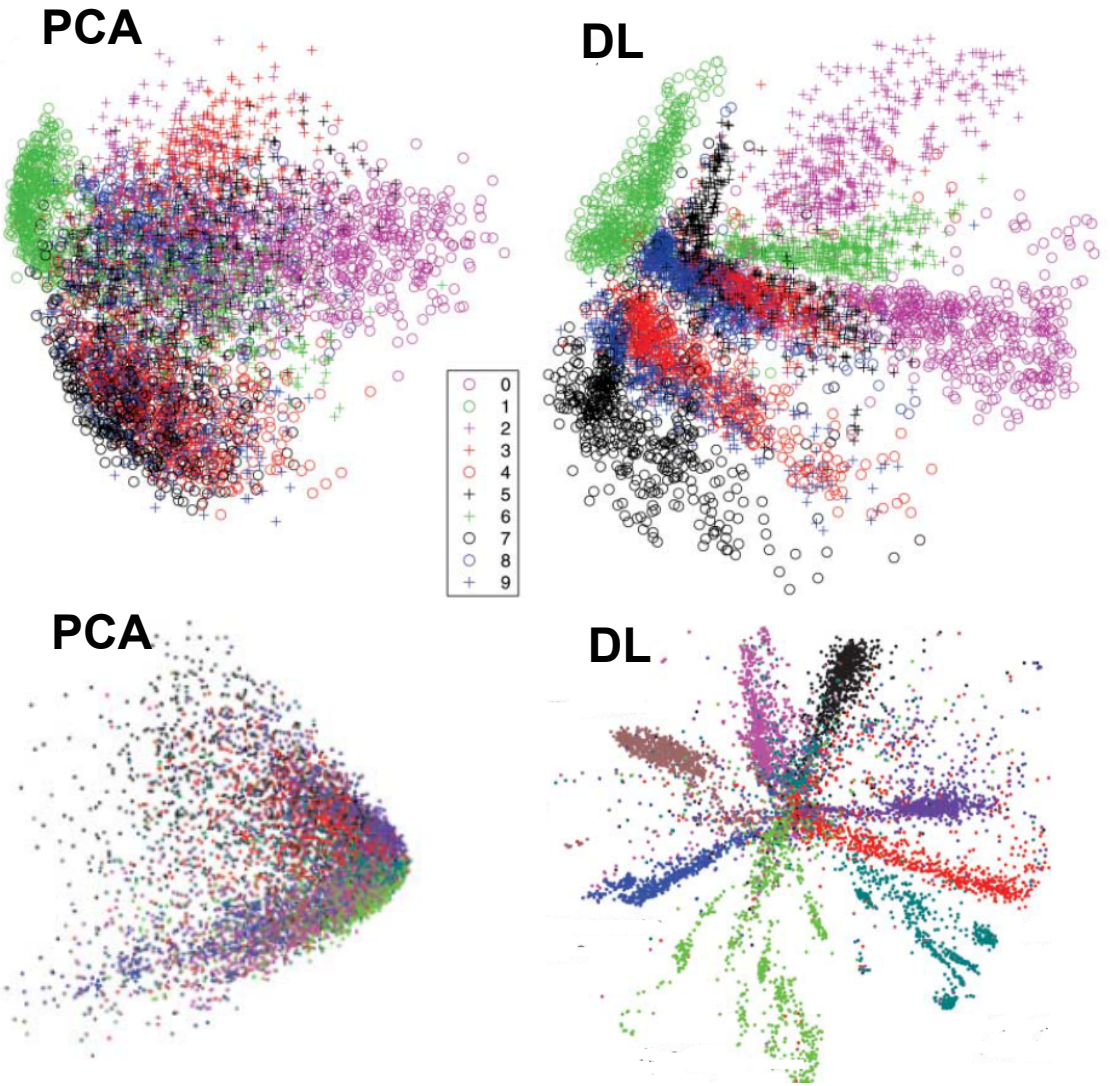
Test compounds with lower reconstruction error are supposed to have more chances to belong to the same activity class as the training compounds



P.V.Karpov, D.I.Osolodkin, I.I.Baskin, V.A.Palyulin, N.S. Zefirov. *Bioorg. Med. Chem. Lett.* **2011**, *21* (*22*), 6728-6731

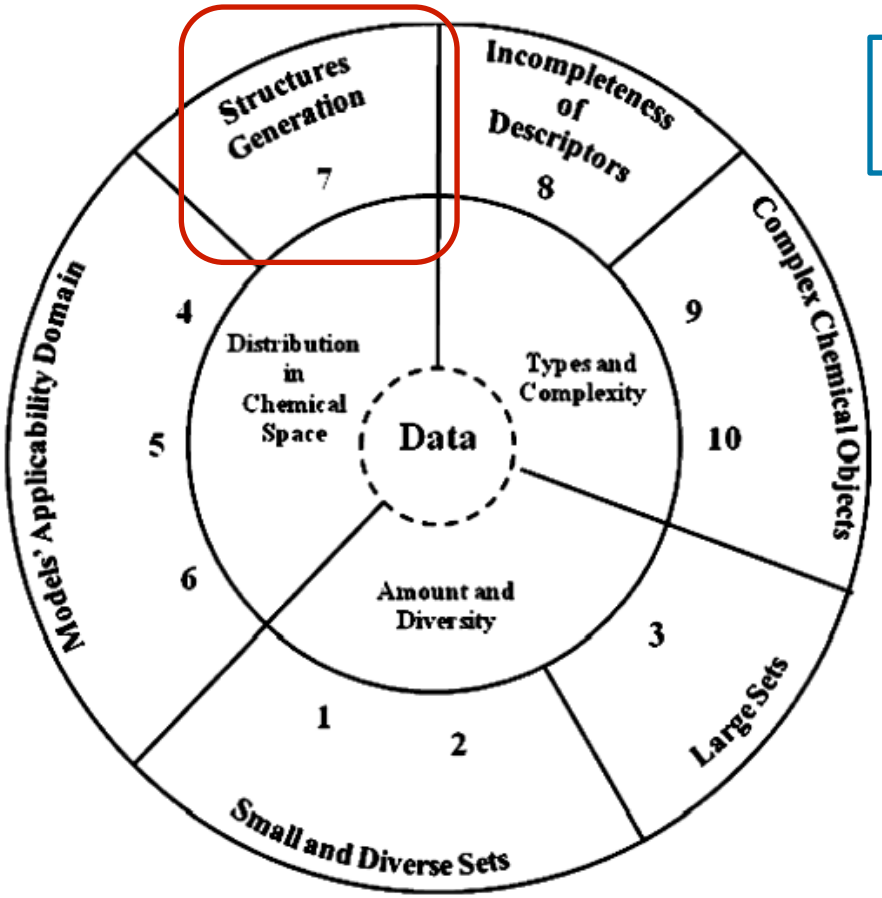
Deep Learning





- G.E.Hinton, R.R.Salakhutdinov, R. R. Science **2006**, *313* (*5786*), 504-507
- Y.Bengio. *Foundations and Trends in Machine Learning* **2009**, *2* (*1*), 1-127

Inverse QSAR



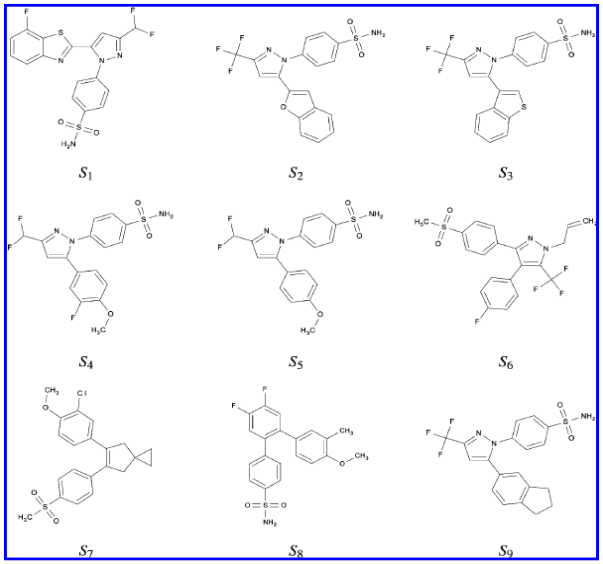
How to generate new chemical structures possessing desired properties?

- Structure generation with filtering through QSAR models
- Combinatorial stochastic optimization utilizing QSAR models
- Solving pre-image problem for kernel-based QSAR models
- Building generative models for graphs

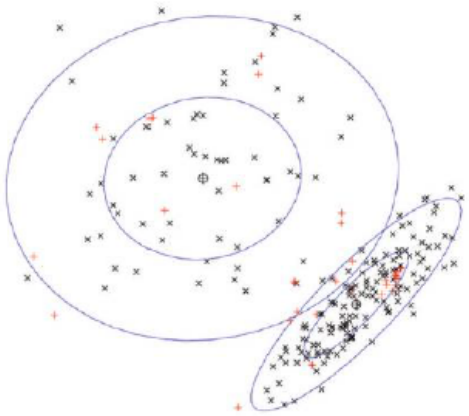
- I.I.Baskin et al. *Dokl. Akad. Nauk SSSR* **1989**, *307* (*3*), 613−617
- Churchwell et al. *J. Mol. Graphics Modell.* **2004**, *22* (*4*), 263−273
- W.Wong, F.A.Burkowski. *J. Cheminf.* **2009**, *1* (*1*), 4.
- D.White, R.C.Wilson. *J. Chem. Inf. Model.* **2010**, *50* (*7*), 1257−1274

Generative Models for Chemical Graphs

Structures for training



COX2 inhibitors

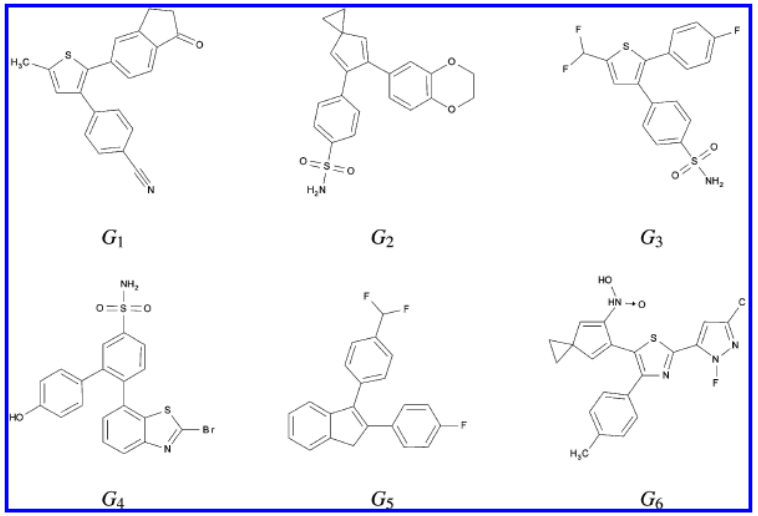


GMM model for P(X|Y)

sampling

Generative models are specified by either joint distribution P(X,Y) or conditional distribution P(X|Y)

$$P(X|Y) = P(X,Y) / P(Y)$$



Generated structures

D.White, R.C.Wilson. *J. Chem. Inf. Model.* **2010**, 50 (7), 1257−1274

JOURNAL OF
CHEMICAL INFORMATION
AND MODELING

Machine Learning Methods for Property Prediction in Chemoinformatics: *Quo Vadis*?

Alexandre Varnek*,† and Igor Baskin†,‡

†Laboratoire d'Infochimie, UMR 7177 CNRS, Université de Strasbourg, 4, rue B. Pascal, Strasbourg 67000, France
‡Department of Chemistry, Moscow State University, Moscow 119991, Russia

Review of existing mathematical approaches potentially useful but rarely or never used in chemoinformatics

A.Varnek, I. Baskin. *J. Chem. Inf. Mod.* **2012**, 52 (6), 1413-1437

Chemoinformatics Tools and the Appropriate Machine Learning Concepts and Methods

Chemoinformatics problem

Machine learning concept

Machine learning method

Implementation in freely available software

	Chemoinformatics task or problem	Machine Learning Concept	Machine Learning method	Implementation in freely available software
1	Increase of the predictive performance of models built on small and diverse data sets	Ensemble learning[191]	Different methods of combining classifiers[192]	meta/Vote (W)
			Bagging[79]	meta/Bagging (W), adabag (R)
			Boosting (classification)[88]	meta/AdaBoostM1 (W), ada, adabag (R)
			Boosting (regression)[91]	meta/AdditiveRegression (W) GAMBoost, mboost (R)
			Stacking[86]	meta/Stacking (W)
			Random subspace[85]	meta/RandomSubSpace (W)
			Random forest[80]	trees/RandomForest (W) randomForest (R)
		Semisupervised and transductive learning[96,193]	TSVM (transductive SVM)[97,194,195]	SVMlight[196]
			SGT (Spectral Graph Transducer)	SGTlight[197]
			SemiL (Semisupervised Learning)[250]	SemiL[198]
			LapSVM (Laplacian SVM),[199] Semisupervised learning based on one-class classification[200] and ensemble learning[201]	
		Inductive knowledge transfer,[201] multitask learning,[153,154] collaborative filtering[224]	Multitask learning using backpropagation neural networks[154]	RSNNS, AMORE, neuralnet, nnet (R); SNNS[202]
			Multitask learning using multitask kernels,[155] Bayesian multitask learning,[203,204] multitask learning using Partial Least Squares (PLS) method,[205] online multitask learning,[206] multitask learning with data editing,[207] semisupervised multitask learning using Dirichlet process,[208] conic programming for multitask learning,[209] multitask learning by multiple kernel learning[210]	
		L_1-Regularized methods	Support Vector Machines(SVM)[6,45,311,312]	functions/SMO (W); kernlab (R); LibSVM,[313] SVMlight[196]
			Ridge regression[314]	functions/LinearRegression (W); Penalized, RXshrink (R)

A.Varnek, I. Baskin. *J. Chem. Inf. Mod.* **2012**, *52* (*6*), 1413-1437

Acknowledgements

- Alexandre Varnek
- Gilles Marcou
- Dragos Horvath
- Nathalie Kireeva

Strasbourg University

- Nelly Zhokhova
- Pavel Karpov
- Dmitry Osolodkin

Lomonosov Moscow State University

- Igor Tetko

Helmholtz Zentrum München