A methodology for Bayesian machine learning and its application to ligand discovery.

Nicos Angelopoulos

n.angelopoulos@ed.ac.uk

School of Biological Sciences

Biochemistry Group

University of Edinburgh, Scotland, UK.

Talk Structure

Principles of Bayesian machine learning.
(Collaborative work with James Cussens, York University.)

Application to classification of affinity data based on high-dimensional descriptor data.

MCMC Overview

Class of sampling algorithms that estimate a posterior distribution.

Markov chain

construct a chain of visited values, $M_1, M_2, \ldots, M_n$, by proposing M_* from M_i , with probability $q(M_*, M_i)$. Use prior knowledge, $p(M_*)$ and relative likelihood of the two values, $p(D|M_*)/p(D|M_i)$ to decide chain construction.

Monte Carlo

Use the chain to approximate the posterior $p(M|D)$.

Bayesian learning with MCMC

Given some data D and a class of statistical models $\mathcal{M}$ ($M \in \mathcal{M}$) that can express relations in the data, use MCMC to approximate normalisation factor in Bayes' theorem

$$p(M|D) = \frac{p(D|M)p(M)}{\sum_M p(D|M)p(M)}$$

$p(M)$ is the prior probability of each model

$p(D|M)$ the likelihood (how well the model fits the data)

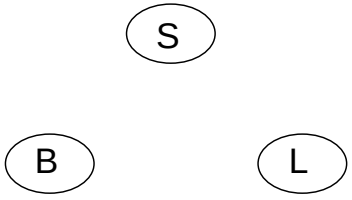
$p(M|D)$ the posterior

Example: Data

	smoker	bronchitis	l_cancer
person 1	y	y	n
person 2	y	n	n
person 3	y	y	y
person 4	n	y	n
person 5	n	n	n

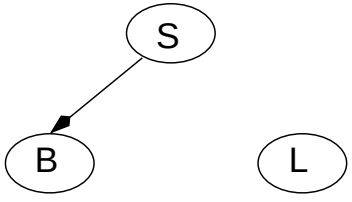
Example: Models

B_1



[b-[],l-[],s-[]]

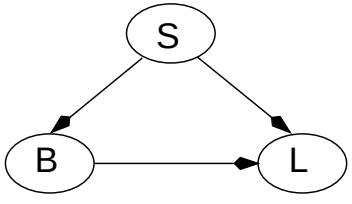
B_2



[b-[s],l-[],s-[]]

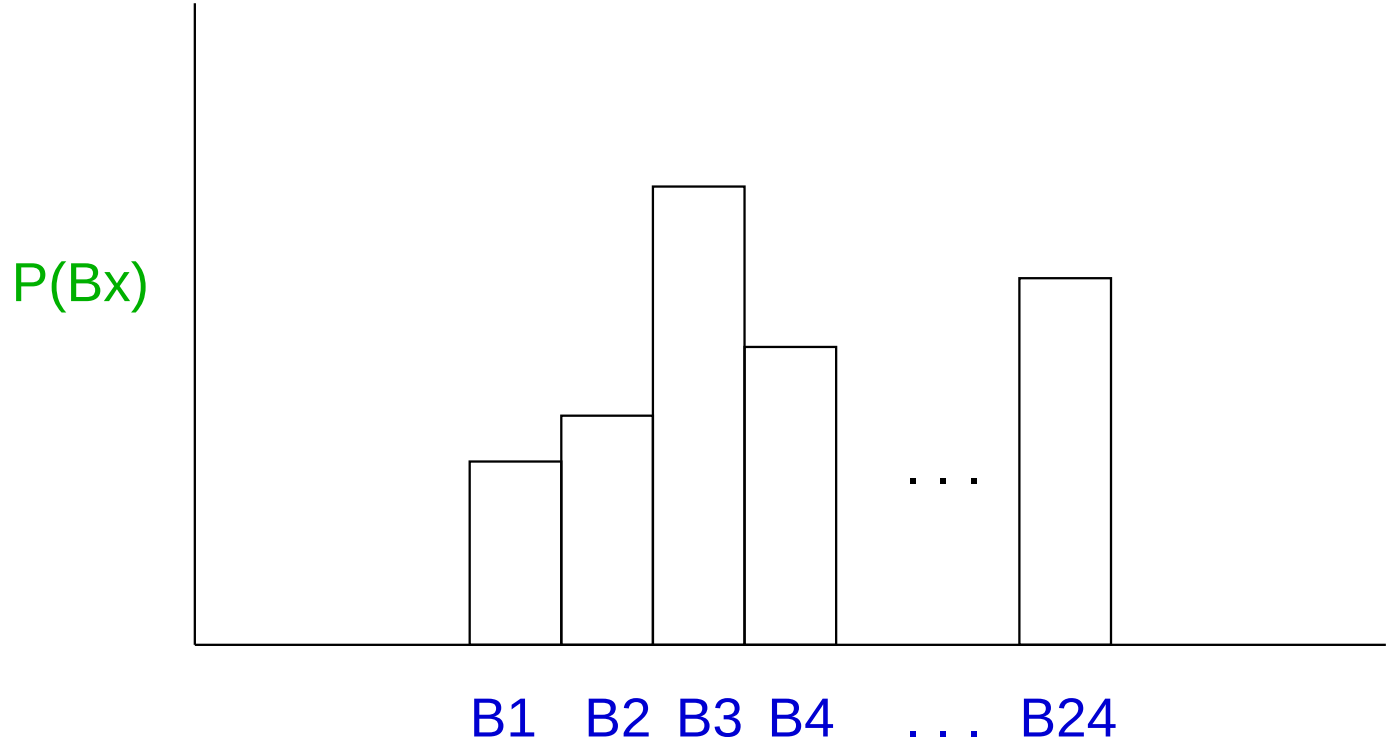
...

B_{24}



[b-[s],l-[b,s],s-[]]

Example: Objective



$$\sum_{B_x} p(B_x) = 1$$

Metropolis-Hastings (M-H) MCMC

0. Set $i = 0$ and find M_0 using the prior.
1. From M_i produce a candidate model M_*. Let the probability of reaching M_* be $q(M_*, M_i)$.
2. Let

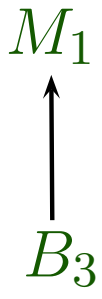
$$\alpha(M_i, M_*) = min\left\{\frac{q(M_*, M_i)P(D|M_*)P(M_*)}{q(M_i, M_*)P(D|M_i)P(M_i)}, 1\right\}$$

$$M_{i+1} = \begin{cases} M_* & \text{with probability } \alpha(M_i, M_*) \\ M_i & \text{with probability } 1 - \alpha(M_i, M_*) \end{cases}$$

3. If i reached limit then terminate, else set $i = i + 1$ and repeat from 1.

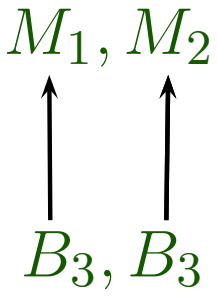
Example: MCMC

Markov Chain:



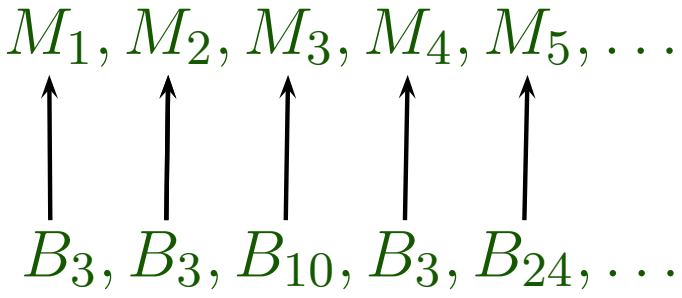
Example: MCMC

Markov Chain:



Example: MCMC

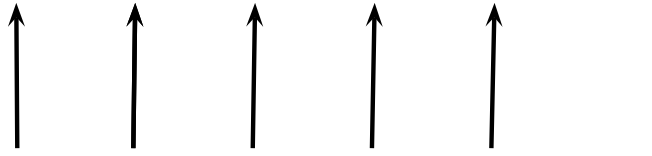
Markov Chain:



Example: MCMC

Markov Chain:

$$M_1, M_2, M_3, M_4, M_5, \ldots$$

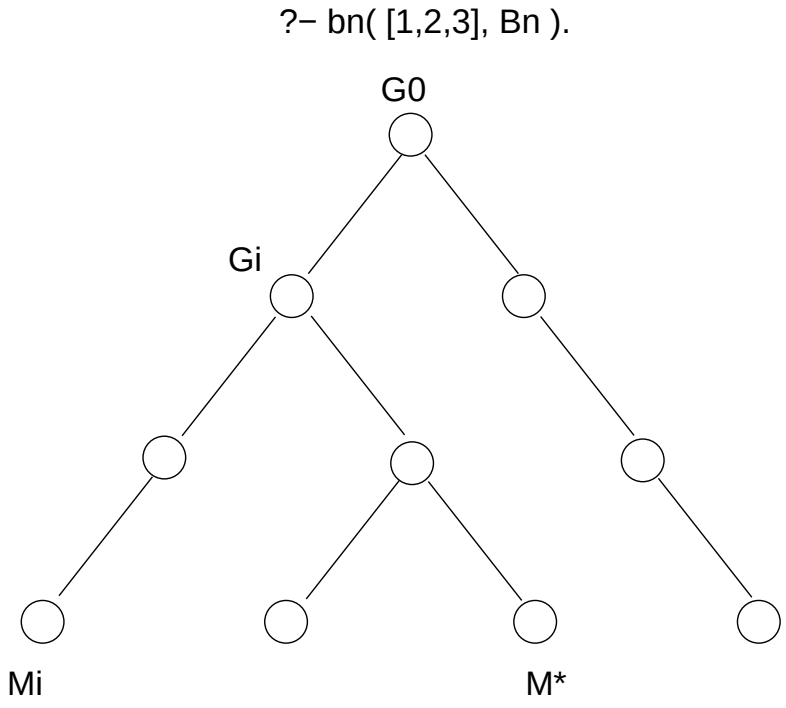


$$B_3, B_3, B_{10}, B_3, B_{24}, \ldots$$

Monte Carlo:

$$p(B_k) = \frac{\#(B_k)}{\sum_{B_x} \#(B_x)}$$

DLP defined model space



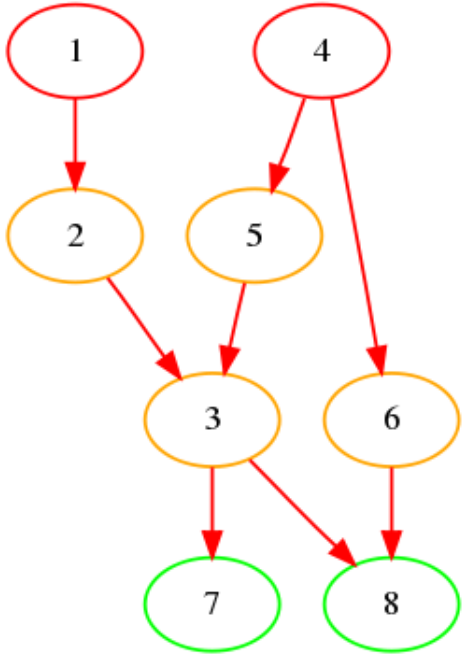
From M_i identify G_i then sample forward to $M_\star$.
$q(M_i, M_\star)$ is the probability of proposing $M_\star$ when M_i is the current model.

BN Prior

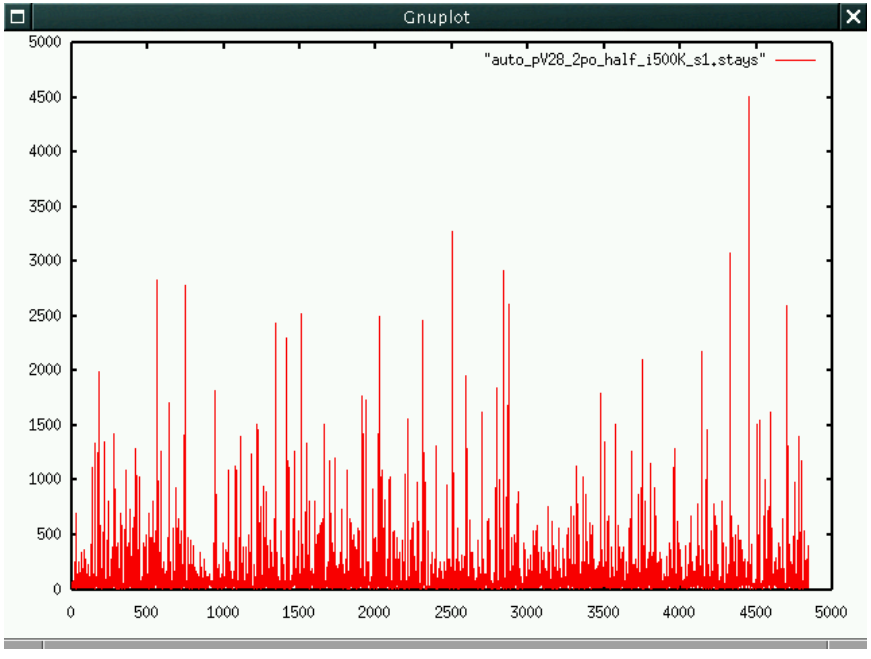
```
bn( OrdNodes, Bn ) :-
    bn( Nodes, [], Bn ).

bn( [], _PotPar, [] ).
bn( [H|T], PotPar, [H-SelParOfH|RemBn] ) :-
    select_parents( PotPar, H, SelParOfH ),
    bn( T, [H|PotPar], RemBn ).

select_parents( [], [] ).
select_parents( [H|T], Pa ) :-
    include_element( H, Pa, RemPa ),
    select_parents( T, TPa ).

1/2 : include_element(  H, [H|TPa], TPa ).
1/2 : include_element( _H,    TPa,  TPa ).
```

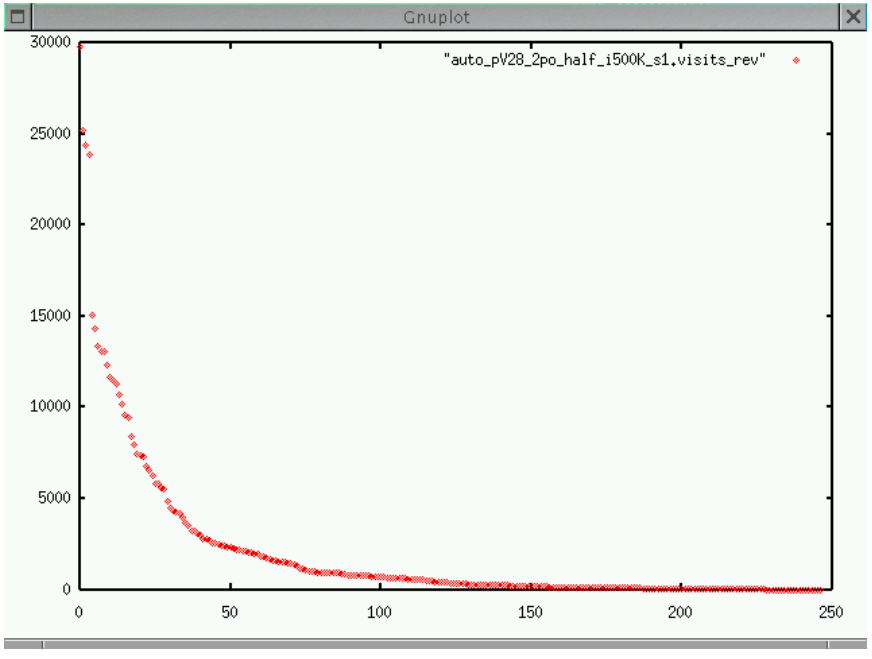
example BN (Asia)

For example ?- bn([1,2,3,4,5,6,7,8], M).
M = [1-[],2-[1],3-[2,5],4-[],5-[4],6-[4],7-[3],8-[3,6]].



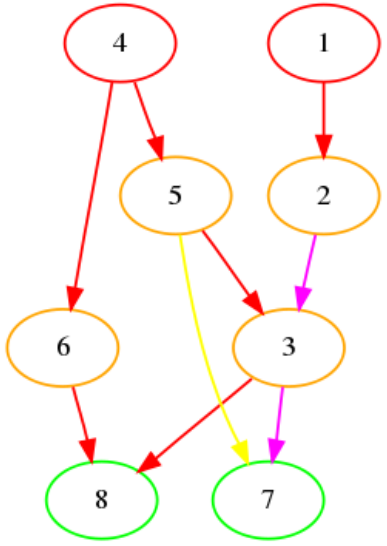
visits and stays





Edges recovery

With topological ordering constraint and a maximum of 2 parents per node, the algorithm recovers most of the BN arcs in 0.5 M iterations. For example for a .99 cut-off we have :



Missing :

- $2 \rightarrow 3$ (.84)
- $3 \rightarrow 7$ (.47)

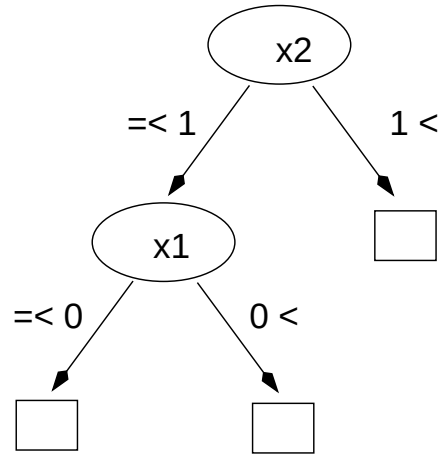
Superfluous :

- $5 \rightarrow 7$

CART priors

$$P_{\text{split}}(\eta) = \alpha(1 + d_\eta)^{-\beta}$$

?- cart(ζ, ξ, A, M).



M=nd(x2,1,nd(x1,0,lf,lf),lf)

(C_0) $cart(\zeta, \xi, A, Cart) :-$

ψ_0 is ζ,

ψ_0: $split(0, A, Cart)$.

(C_1) ψ_D: $split(D, \zeta, \xi, A_B, nd(F, Val, L, R)) :-$

ψ_{D+1} is $\zeta * (1 + D)^{-\xi}$,

D_1 is $D + 1$,

$r\_select(F, Val, A_B, L_B, R_B)$,

ψ_{D+1}: $split(D_1, \zeta, \xi, L_B, L)$,

ψ_{D+1}: $split(D_1, \zeta, \xi, R_B, R)$.

(C_2) $1 - \psi_D$: $split(D, \zeta, \xi, A_B, lf)$.

Virtual Screening experiment

Objective: improve chances of discovering binding molecules based on examples from screened chemical libraries.

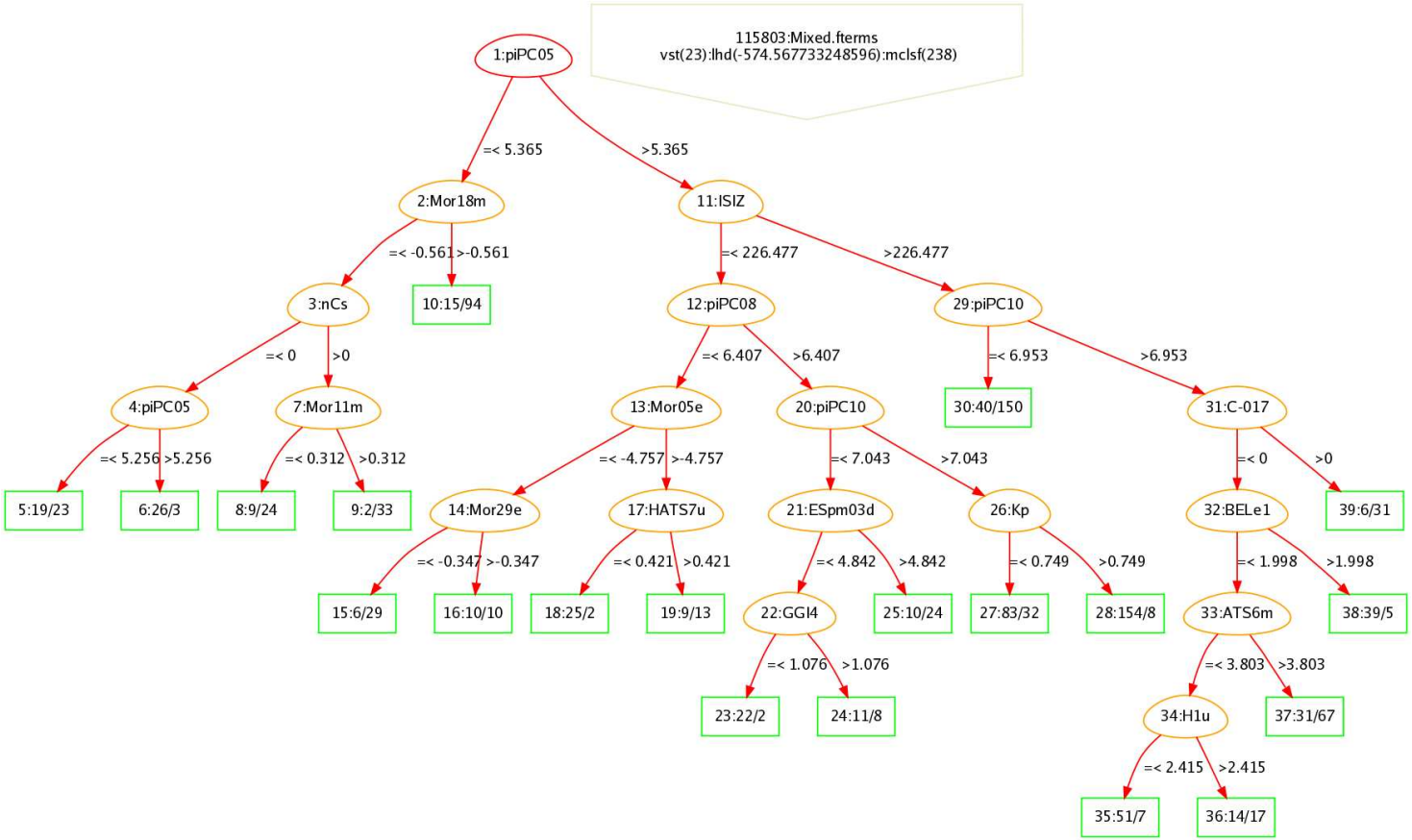
Pyruvate kinase affinity data. 582 Active and 582 Inactive compounds selected. The program Dragon was used to produce 1500 property descriptors for each molecule of used normaly 1100 were used.

Compared to Feed Forward Neural Networks by spliting the data into ten train/test segments.

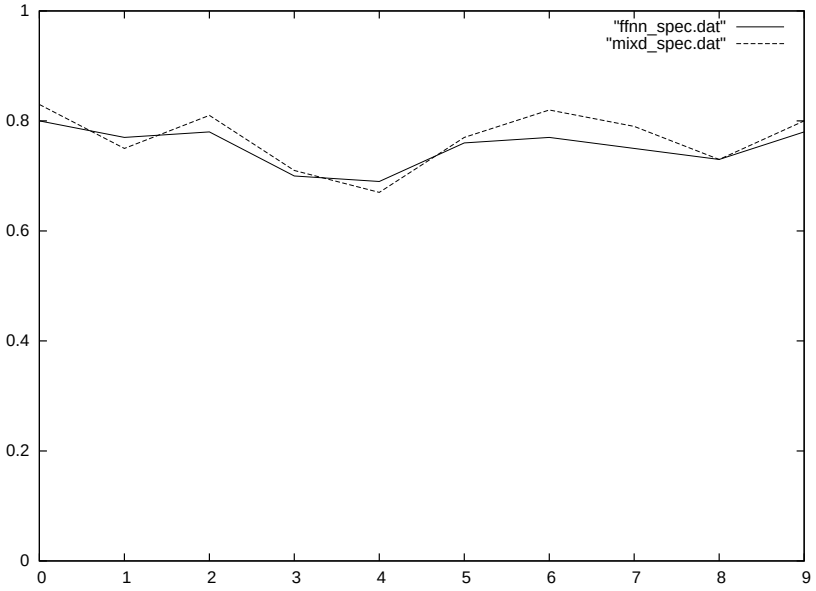
First segment results

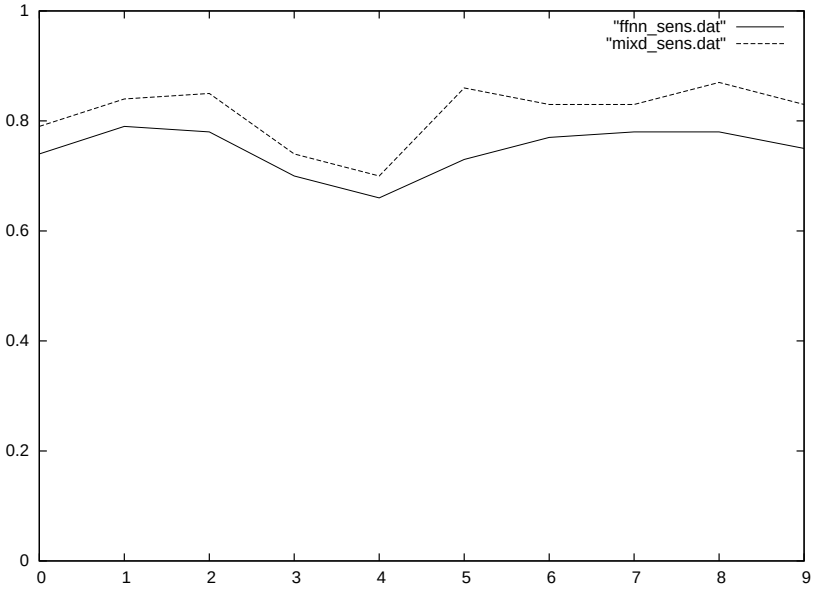
	true +ve		true -ve			
	mean	std err	mean	std err	Sensitivity	Specificity
MH	45	1.25	37.67	1.78	0.6888	0.7434
MH-Aw2	45.7	0.720	39.33	1.90	0.7099	0.7617
MH-T3	45.3	1.44	43.67	1.52	0.7596	0.7747
FFNNE	45	0	42	0.816	0.7377	0.7636
FFNN	47.67	0.272	41.67	0.720	0.7448	0.8013
Mixed	48.3	0.272	45.3	0.544	0.7922	0.8255
Mixed-Aw2	49	0.471	45.33	0.272	0.7945	0.8343
Mixed-T3	49.33	0.728	46.33	0.272	0.8086	0.8423

highest likelihood model



Ten-fold validation

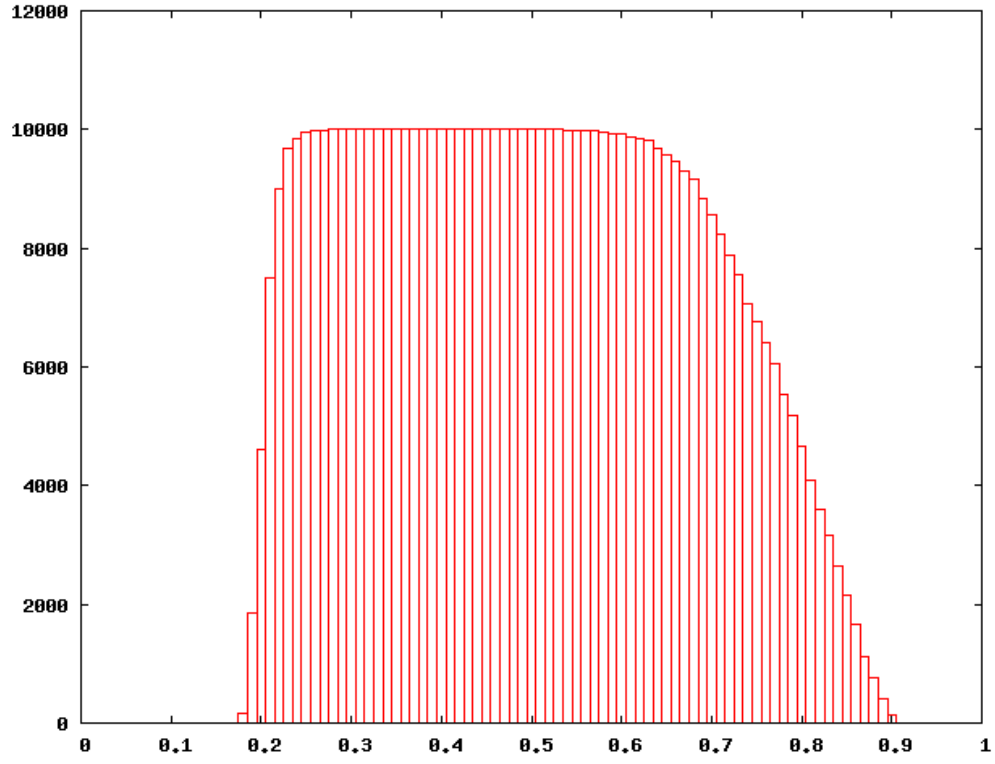




$$Specificity = \frac{T^-}{T^- + F^+}$$

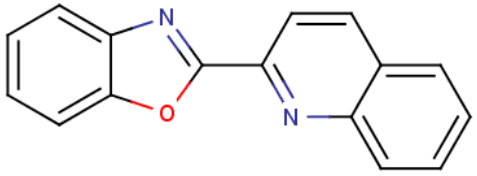
$$Sensitivity = \frac{T^+}{T^+ + F^-}$$

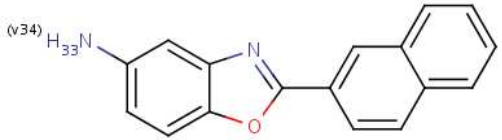
EDULISS DB Classification



http://eduliss.bch.ed.ac.uk/eduliss/

higest probability hits





bottom line

advantages for DB classification:

- posterior provides a rank
- allows further analysis and feature weighing
- can classify molecules with missing values

future work

- test top hits in wet-lab
- apply to further datasets
- improve Chain mixing