Neural Network Applications

Peter Tiňo

P.Tino@cs.bham.ac.uk

School of Computer Science
University of Birmingham

Outline

- 'On the Importance of Being ...
  ... Able to Learn'
- Brain, neurons and all that ...
- Formal (artificial) neurons and networks of artificial neurons
- Application of NN in drug development (Pfizer)

Why learning?

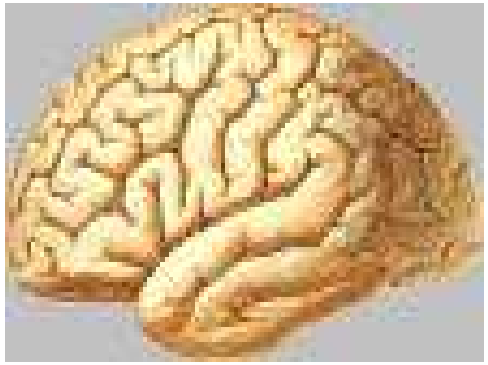
- Simple organisms vs. complex species
- Functionality can be wired up, but in complex organisms this is infeasible. We need a mechanism to learn from experience.
- So what is wired up is a mechanism to learn from experience
- Ability to learn can be good even for computers!

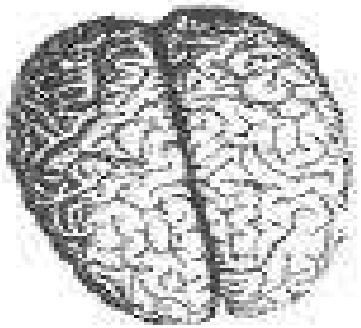
Human Brain

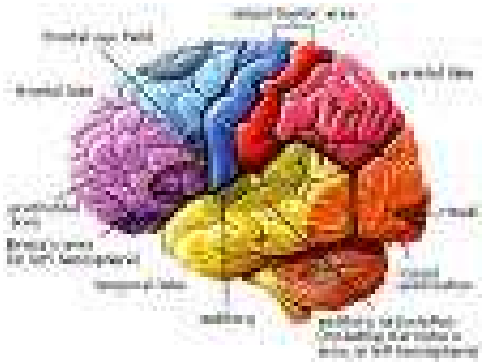
- Brain – highly complex, non-linear and parallel information processing system (computer)
- Composed of neurons
- On certain tasks it is much faster than supercomputers of today.
  E.g. recognise a familiar face in an unfamiliar scene in 100-200 ms.
- At birth brain is already highly structured, but dramatic development continues for the first 2 years.
- Plasticity of the neural network constituents

Human Brain

- Typically, neurons (10^{-3} sec range [ms]) are 5-6 orders of magnitude slower than silicon logic gates (10^{-9} sec range [ns])
- Huge number of interconnected neurons.
  10 billion neurons in the human cortex.
  60 trillion synapses (connections).
- 10 orders of magnitude more energy efficient than computer.

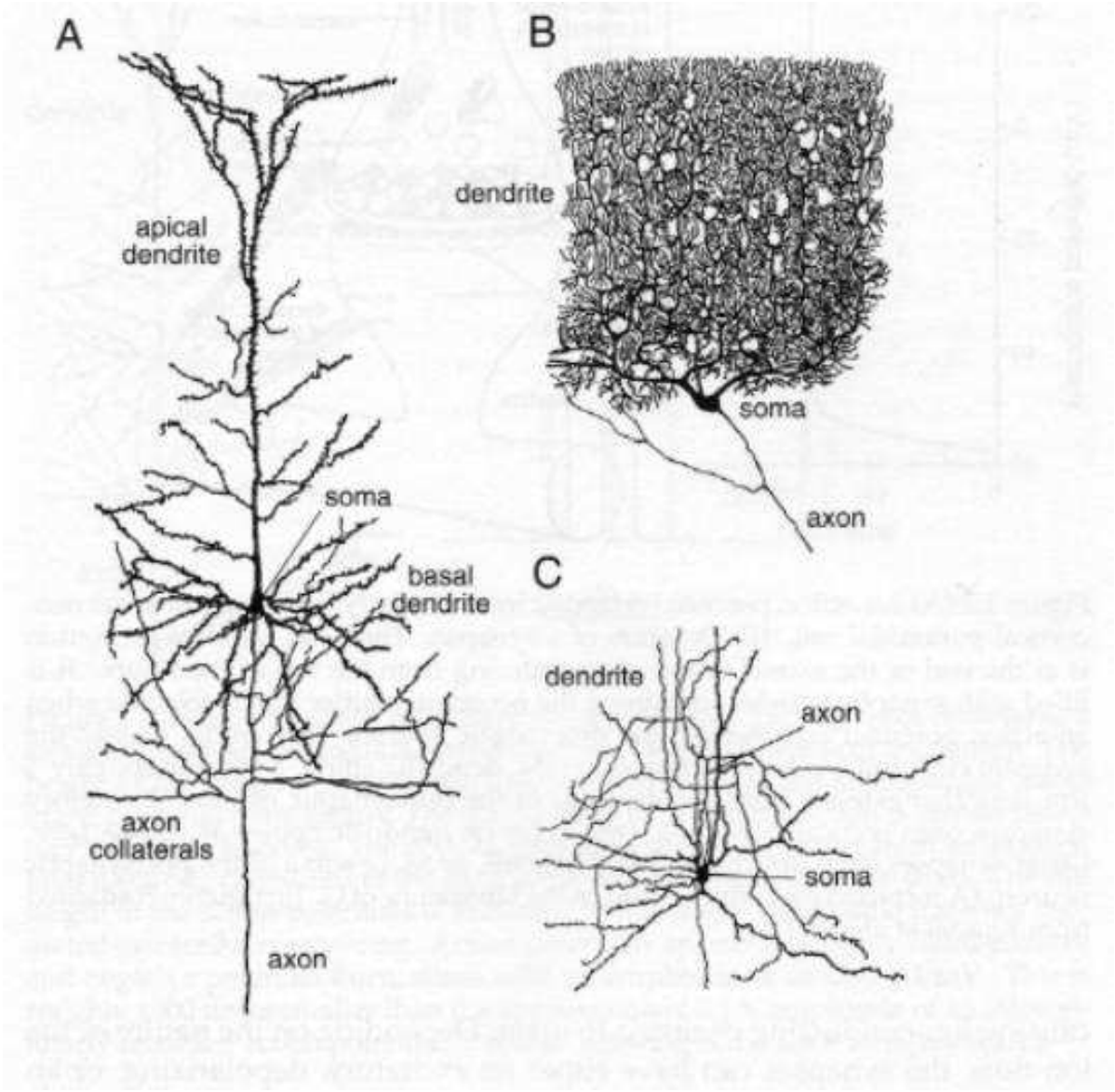






Plasticity

- Plasticity permits nervous system to adapt to its environment.
  Two mechanisms:
  (1) creation of new synaptic connections between neurons
  (2) modification of existing synapses
- Axons – transmission lines.
  Smooth surface, few branches, long.
- Dendrites – receptive zones.
  Irregular surface, many branches, short.
- Majority of neurons encode their outputs as a series of spikes (voltage pulses).



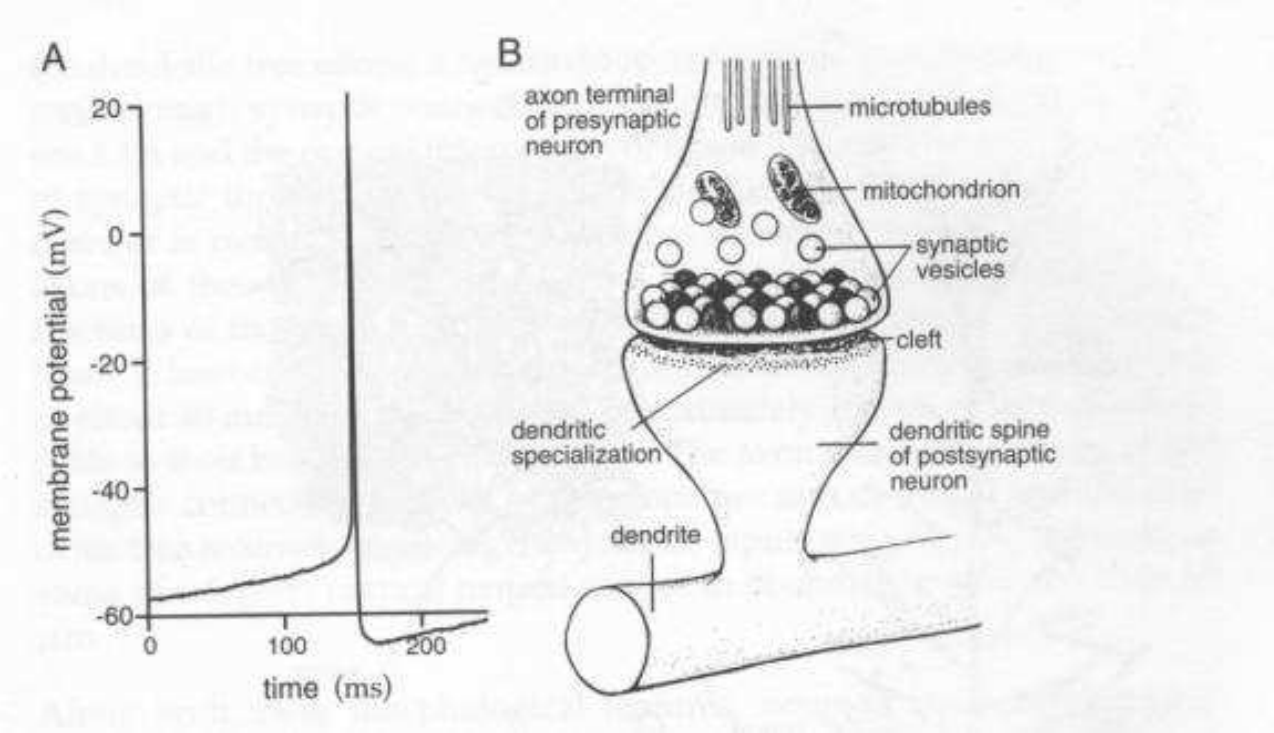
Extensive synaptic connectivity is a hallmark of neural circuitry.

- Figure A:
  A cortical pyramidal cell.
  These are the primary excitatory neurons.
  Branch locally, sending axon collaterals to synapse with nearby neurons.
  Also project more distally to conduct signals to other parts of the brain.

- Figure B:
  A Purkinje cell.
  Receive 100,000 of synaptic inputs.
  Axons transmit the output of the cerebellar cortex.

- Figure C:
  A stellate cell.

Synapse



Action potential recorded from a rat neocortical pyramidal cell.

Neuron - Synapse

- Synapses mediate the interaction between neurons.
  Presynaptic process liberates a transmitter substance.
  The substance diffuses across the synaptic junction between neurons and then acts on a postsynaptic process.
- Synapse converts presynaptic electrical signal into a chemical signal and then back into a postsynaptic electric signal.
- Synapse can be viewed as a simple connection that can impose excitation or inhibition, but not both on a receptive neuron.

When does a neuron 'fire'?

- stimulation of a neuron requires usually either
  (1) repetition of impulses in time at the same synapse (temporal summation), or
  (2) simultaneous arrival of impulses at a sufficient number of adjacent synapses to make the "density" of excitation high enough at some region of the neuron.
- When synaptic excitation takes place, the passage of the impulse across the synapse consumes time (synaptic delay).
- The arrival of impulses at synapses may have the opposite effect, i.e., it may render the element less excitable to other stimuli (inhibition).

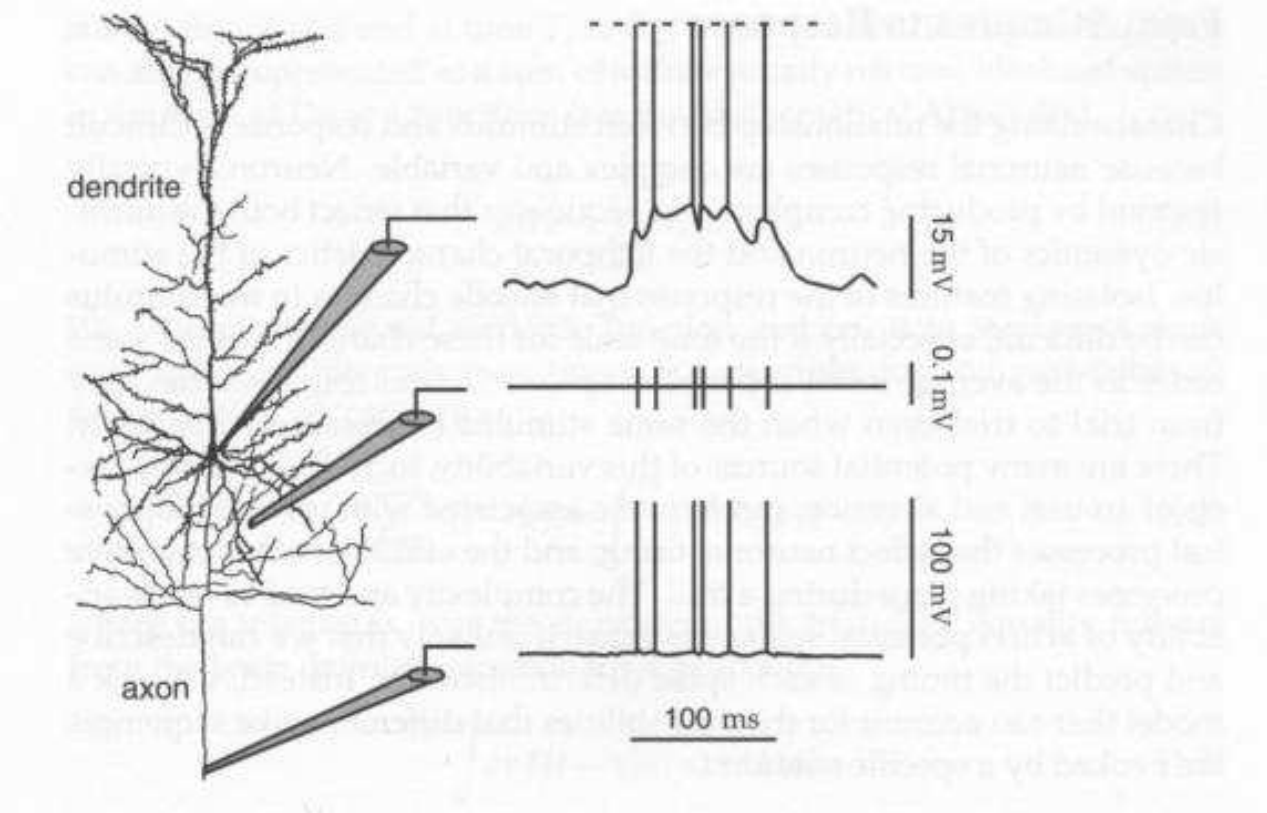
Measuring neural activations

- Membrane potentials are measured intracellularly by connecting a hollow glass electrode filled with a conducting electrolyte to a neuron, and comparing the potential it records with that of a reference electrode placed in the extracellular medium.
- Intracellular recordings are made e.g. with sharp electrodes inserted through the membrane into the cell.
- Intracellular recording is more commonly used for in vitro preparations, such as slices of neural tissue.

Measuring neural activations

- In extracellular recording the electrode is placed near a neuron, but it does not penetrate the cell membrane. Such recordings can reveal action potentials fired by a neuron, but not its subthreshold membrane potentials.
- Extracellular recordings are typically used for in vivo experiments, especially those involving behaving animals. Intracellular recordings are sometimes made in vivo, but this is difficult to do.

Communicating through signals



simulated recordings from a neuron.

- Top trace:

  Recording from an intracellular electrode connected to the soma of the neuron.

  The recording shows rapid spikes riding on top of a more slowly varying subthreshold potential.

- Bottom trace:

  Recording from an intracellular electrode connected to the axon some distance away from the soma.

  The subthreshold membrane potential waveform, apparent in the soma recording, is completely absent on the axon due to attenuation, while the action potential sequence in the two recordings is the same.

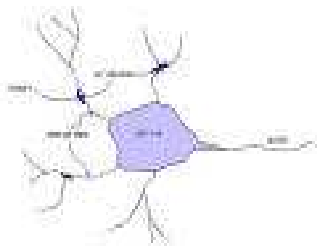
- Middle trace:

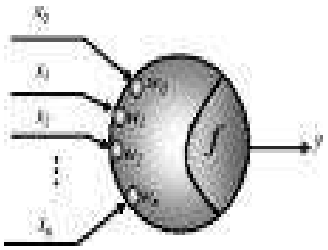
  A simulated extracellular recording.

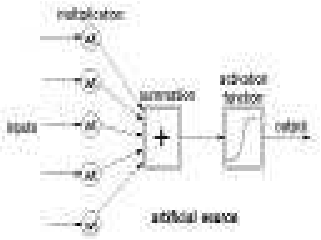
In the end it is down to spikes...

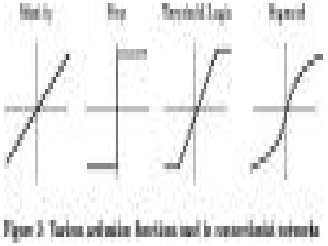
- The previous example illustrates an important point: Spikes, but not subthreshold potentials, propagate regeneratively down the axons.

Formal neurons



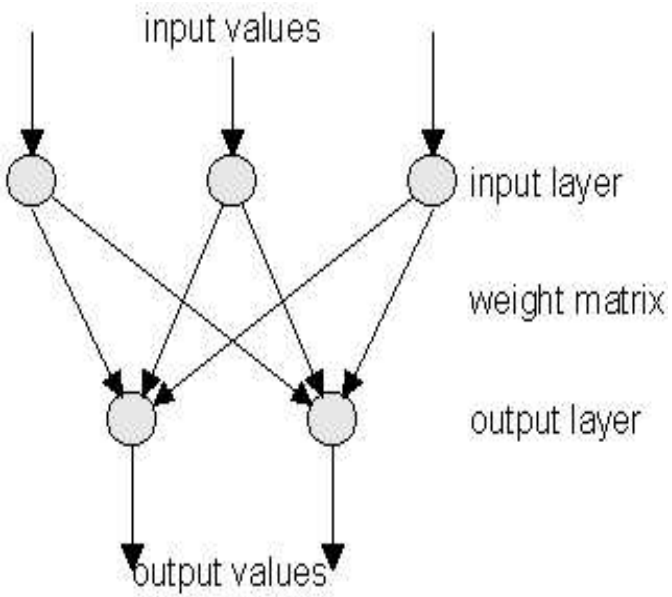


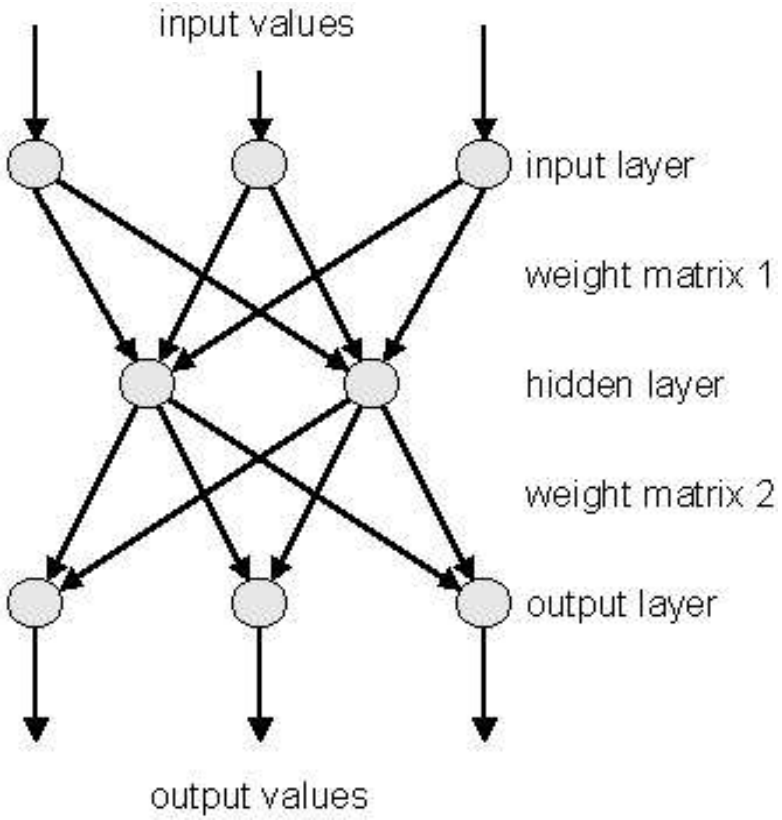




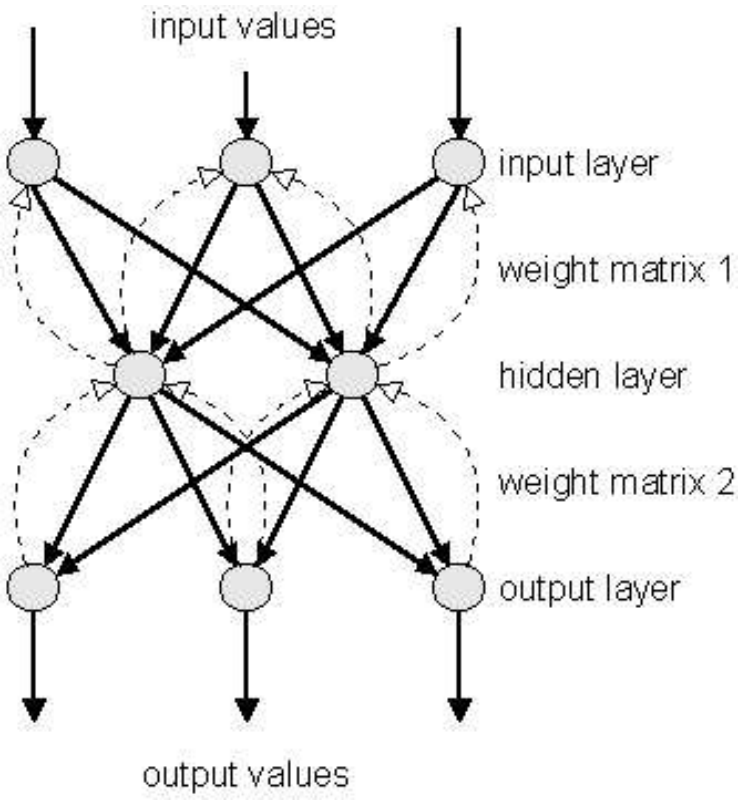
- Rate coding in time (single neuron)
- Rate coding in space (pools of neurons)

Networks of formal neurons





Learning in neural networks



Quantitative Structure-Activity Relationships (QSAR)

- Going from molecular structure to the activity on other molecular compounds can be a tricky business.
- Molecular structure: 3-D arrangement of atoms (of various types: H, O, N, ...) and bonds between them (of various types)
- The way the activity levels are coded in the 3-D structure is
  - not well-understood –
    hence the need to use a learning system
  - very complicated –
    hence the need to use a sophisticated learning system

Partition coefficient P

- Majority of pharmaceutical agents must cross a biological membrane to reach their site of action and to be available in a cellular environment.
- Lipophilicity of the 'drug' molecule has a major impact upon its distribution and biological action.
- Quantitative measures of lipophilicity are very important in the development of drug molecules.
- partition coefficient of a molecule is the ratio of its solubility in n-octanol to its solubility in water
- Logarithm of this quantity, LogP, is a well established measure of a compound's lipophilicity.

Determine $LogP$ without actually measuring it ...

- In practice, measurement of LogP for large numbers of compounds is costly and time consuming
- Computational methods are employed to estimate or predict values where possible
- Even more importantly, it is valuable to have an estimate of lipophilicity before synthesising novel compounds, and this can be only be done using a predictive model.
- Numerous methods for calculating LogP exist, mainly characterised as substructural and whole molecular approaches.

State-of-the-art commercial programs

- ClogP,AlogP, MlogP, ...
- Often based on complex molecular representations.
- Quite expensive.

Molecular descriptors

- 14 simple molecular descriptors based on a two-dimensional representation of the molecular structure (Pfizer)
- Descriptors 1–10: InterAction Fingerprints (IAFs).
  IAFs are the average counts for non-covalent interactions (strong, medium, weak hydrogen bonds, Van der Waals and pi-interactions) around individual atom types, summed up over the whole molecule.
- Descriptor 11: Sum of volumes of Voronoi polyhedra (used to determine the IAFs).
  Used as a measure of size.
- Descriptor 12: Halogen counts for fluorine, chlorine and bromine.

Datasets

- Dataset I:
  6912 compounds together with their LogP values that is freely available on the Internet. 10% of the set was used as a test set and, when needed, another 10% was set apart as a validation set for model selection. The remaining data was used as a training set.

- Dataset I+II:
  Dataset I augumented by Dataset II, a new set of 226 compounds whose LogP values were measured at Pfizer. Dataset II served as a completely blind test set for models trained on the whole of Dataset I. When needed, 10% of Dataset I was set apart as a validation set for model selection.

Results - Dataset I

	Dataset I		
model class	MSE	ION	selected model
Naive	2.690	0	Naive
LR	0.799	70.3	LR
MLP	0.641	76.2	$N_{hid} = 15$
MLP-ARD	0.611	77.3	$N_{hid} = 15$
HME	0.658	75.5	BF=2, D=3

Results - Dataset I+II

	Dataset I+II		
model class	MSE	ION	selected model
Naive	4.075	0	Naive
LR	1.190	70.8	LR
MLP	1.075	73.6	$N_{hid} = 15$
MLP-ARD	1.119	72.5	$N_{hid} = 15$
HME	0.953	76.6	BF=2, D=3

Results - State of art ...

	Dataset I		Dataset I+II	
method	MSE	ION	MSE	ION
ClogP	0.175	93.5	1.023	74.9
AlogP	0.550	79.5	1.164	71.4
MlogP	2.109	21.6	1.088	73.3
IAF	1.318	51.0	1.077	73.6

Conclusions

- Simple compound representations can lead to results comparable to those of commercial state-of-art products operating on more complicated molecular representations.
- Carefully constructed committees of NNs operating on the novel representation achieve results comparable with those of more complex and expensive state-of-art products.
- Generalisation performance on previously unseen compounds is precisely the situation of most value to working chemists.