Neuronale Netze (SS 2002), 19.6.

And the remaining parts of COLT theory

- Some examples for the VC dimension:
  - VC(all possible functions) $= \infty$
  - VC(finite function class $\mathcal{F}$) $\leq \log_2 |\mathcal{F}|$
  - VC(H $\circ$ polynomials in $\mathbb{R}$ with degree at most d) $\leq d+1$
  - VC(perceptron with input dimension n) $= n+1$
- The VC-dimension allows to estimate the size of the function class:

  for every probability measure P it holds

  $$N(\epsilon, \mathcal{F}, d_P) \leq \left(\frac{2\,\mathrm{e}}{\epsilon} \ln \frac{2\,\mathrm{e}}{\epsilon}\right)^{\mathrm{VC}}$$

  Conversely, if VC $= \infty$, then we can find for every number m a probability measure P with $N(\epsilon, \mathcal{F}, d_P) \geq m$.

  $\Rightarrow$ distribution independent PAC learnability and the distribution independent UCED property are both equivalent to a finite VC-dimension. In this case, every learning algorithm with small error will do and $\sim$ VC examples are sufficient for every learning algorithm for valid generalization.
- Some more VC-dimensions:
  - VC(perceptron networks with W weights) $= \Theta(W \log W)$
  - one can find bounds for the VC-dimensions of neural networks with piecewise polynomial activation functions
  - VC(sigmoidal networks with N neurons and W possibly shared weights) $= O(W^2 N^2)$ and $= \Omega(W^2)$

    (... whoever solves the gap will become famous ;-)
- Finally a particularly important alternative estimation for the VC-dimension of perceptrons ...

Note: the VC-dimension of an SVM equals the VC-dimension of the (trainable) perceptron in the feature space where data are implicitely mapped to with the (fixed) nonlinear mapping. The dimensionality is usually high dimensional, possibly infinite dimensional.

Alternative: Assume SVM has a margin γ and the points are absolutely bounded by R, then

$$\text{VC} \leq \frac{R^2}{\gamma^2}$$

$\rightarrow$ the SVM tries to optimize the possibly high or infinite dimensional VC dimension DURING training by maximizing the margin!!!

(In technical terms: 'minimize $|\vec{w}|^2$' is the same as 'maximize the margin' which is the same as 'minimize the VC-dimension' for SVMs.)