Neuronale Netze (SS 2002), 15.4.

Perceptron convergence theorem:

If a training set is linearly separable, a solution is found in a finite number of steps by the perceptron learning algorithm.

Proof idea: $\vec{w}$:= solution with $|\vec{w}^t\vec{x}^i| \geq 1$, $\vec{w}_k$:= perceptron after step k, then

(1) $\vec{w}^t\vec{w}_k \geq \vec{w}^t\vec{w}_0 + k$, (2) $\|\vec{w}_k\|^2 \leq \|\vec{w}_0\|^2 + k \cdot \max^2$

This can hold for only a finite number of steps k!

Questions:

- How long does it take?
  - in-principle bound for start in $\vec{0}$: $k = \|\vec{w}\|^2 \cdot \max^2$,
    thereby, $\|\vec{w}\|$ can be estimated in principle from the patterns
    ($\rightarrow$ set of linear inequalities)
  - there exist problems where perceptron learning takes an exponential number of time steps
  - substitute perceptron learning by polynomial linear programming
    ($\rightarrow$ Karmakar algorithm)
- Does perceptron learning generalize to unseen examples?
  - Yes, can be formalized and proved in statistical learning theory.
    ($\rightarrow$ We'll get this later ...)
- Is 'linearly separable' sufficient?
  - No, the number of linearly separable functions $\{0,1\}^n \rightarrow \{0,1\}$ is only of order 2^{n^2} instead of 2^{2^n}!
    ($\rightarrow$ statistical learning theory ...)
  - Even linearly separable data is usually subject to noise.
- What happens if we train a perceptron for a training set which is not linearly separable? (Well, let's see ...)