Machine Learning 1
Lecture 6.5 - Supervised Learning
Classification - Discriminative Models - The Perceptron

Erik Bekkers

(Bishop 4.1.7)
The Perceptron Algorithm

- Input: \( \mathbf{x} \in \mathbb{R}^D \)
- Targets: \( t \in \{ C_1, C_2 \} \rightarrow t \in \{-1, 1\} \) 2 classes
- Prediction: \( y(x) = f(w^T \phi(x)) \quad f(a) = \begin{cases} 1, & a \geq 0 \\ -1, & a < 0 \end{cases} \)
- Class decisions: assign \( \mathbf{x} \) to class \( C_1 \) if …… \( \mathbf{w}^T \phi_n > 0 \) (and \( C_{-1} \) if \( \mathbf{w}^T \phi_n < 0 \))
- For correct classification: find \( \mathbf{w} \) such that for all \((x_n, t_n)\):
  \[ \mathbf{w}^T \phi_n t_n \geq 0 \]
- Perceptron criterion: \( E_P(w) = - \sum_{n \in \mathcal{M}} w^T \phi(x_n) t_n \)
  \( \mathcal{M} : \{ n : \mathbf{w}^T \phi_n t_n < 0 \} \)
Perceptron: Stochastic Gradient Descent

- \( E_P(w) = -\sum_{n \in M} w^T \phi(x_n) t_n \)
  \[ = \sum_{n \in M} E_n(w) \]

- Stochastic Gradient Descent (SGD).
  For each misclassified \( x_n \):

  \[ w^{(\tau+1)} = w^{(\tau)} - \eta \nabla E_n(w) \]
  \[ = w^{(\tau)} + \eta (\phi_n b_n) \]

- If \( X \) is linearly separable, then perceptron SGD will converge

**Figure:** for \( x_n \) in \( C_1 \): add \( \phi(x_n) \) to \( w \), for \( x_n \) in \( C_2 \): subtract \( \phi(x_n) \) from \( w \). SGD for perceptron criterion (Bishop 4.7)
Problems: Perceptron

- Perceptron only works for 2 classes

- There might be many solutions depending on the initialization of $\mathbf{w}$ and on the order in which data is presented in SGD

- If dataset is not linearly separable, the perceptron algorithm will not converge.

- Based on linear combination of fixed basis functions.