Recognition Labs – Nonlinear Perceptron

Vojtv ech Franc, translated by Petr Doubek
Center for Machine Perception, FEE CTU Prague

November 5 2009

1 Introduction

The goal of this exercise is to extend the perceptron learning algorithm to nonlinear classifer, i.e. classifier with nonlinear discriminative function. We will demonstrate it on learning a quadratic discriminative function in a two-dimensional feature space.

2 Lifting the Feature Space

The perceptron algorithm optimizes parameters of a linear classifier, so that the classifier has zero error on given training data $\{(\vec{x}_1, y_1), \ldots, (\vec{x}_m, y_m)\}$. The linear classifier decides based on the sign of linear discriminative function

$$f(\vec{x}) = \langle \vec{w}, \vec{x} \rangle + b = \sum_{i=1}^{n} w_i x_i + b \,, \tag{1}$$

where vector $\vec{w} \in \mathbb{R}^n$ and scalar $b \in \mathbb{R}$ are parameters and $\vec{x} \in \mathbb{R}^n$ is a feature vector of the object to be classified. Note that the discriminative function (1) is linear in both parameters $(\vec{w}, b)$ as well as in the feature vector $\vec{x}$. By nonlinear classifier we mean nonlinear with respect to the feature vector $\vec{x}$. For optimizing the parameters of nonlinear classifier by perceptron, we will use *lifting of feature space.* The idea of lifting is as follows: (i) project training samples from input feature space $\mathcal{X} \subseteq \mathbb{R}^n$ into new (lifted) space $\mathcal{Z} \subseteq \mathbb{R}^N$, where nonlinear functions project as linear; (ii) use perceptron algorithm in lifted space $\mathcal{Z}$. Found parameters determine a linear classifier in the space $\mathcal{Z}$

and a nonlinear classifier in the input space $\mathcal{X}$. The function which maps the input feature space to the new lifted space is denoted as $\phi\colon \mathcal{X} \to \mathcal{Z}$. Let us focus on one specific case of learning a quadratic discriminative function: the input feature space $\mathcal{X} \subseteq \mathbb{R}^n$ is two-dimensional ($n = 2$) and the quadratic discriminative function has form

$$q(\vec{x}) = w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_1 \cdot x_2 + w_4 \cdot x_1^2 + w_5 \cdot x_2^2 + b = \sum_{i=1}^{5} w_i \phi_i(\vec{x}) + b\,. \quad (2)$$

By comparing the linear (1) and the quadratic discriminative function (2), we can derive mapping function ϕ

$$\vec{z} = \phi(\vec{x}) \quad \text{as} \quad \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \end{bmatrix} = \begin{bmatrix} \phi_1(\vec{x}) \\ \phi_2(\vec{x}) \\ \phi_3(\vec{x}) \\ \phi_4(\vec{x}) \\ \phi_5(\vec{x}) \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_1^2 \\ x_1 \cdot x_2 \\ x_2^2 \end{bmatrix}. \quad (3)$$

The lifted feature space $\mathcal{Z}$ is 5-dimensional. Entire extension of the linear perceptron algorithm to algorithm searching quadratic discriminative function is done as follows:

1. Transform the training data $\{(\vec{x}_1, y_1), \ldots, (\vec{x}_m, y_m)\}$ into lifted feature space as $\{(\vec{z}_1, y_1), \ldots, (\vec{z}_m, y_m)\}$ using the mapping function (3).
2. Use perceptron algorithm to learn a linear classifier on the lifted training set $\{(\vec{z}_1, y_1), \ldots, (\vec{z}_m, y_m)\}$.
3. Substitute the found parameters $(\vec{w}, b)$ into the quadratic discriminative function (2).