

Forschungsmodul: Komplexe Systeme

Bericht zur Vorlesung vom 15. November 2007 von Jan-Philip Gehrcke

Künstliche neuronale Netzwerke - Lernalgorithmen

Hebbsche Kopplungen

Hat man eine linear separable Dichotomie gegeben, liegt es nahe ein Perzeptron zu konstruieren, welches genau diese Klassifikation ausführt. Gegeben ist also ein linear separabler Satz von Mustern $\vec{x}^{(\mu)}$ mitsamt seiner "Ergebnisse" S_μ :

$$\{\vec{x}^{(\mu)}, S_\mu\}, \quad \mu = 1, \dots, P$$

Die Aufgabe ist es nun, den die Trenn-Hyperebene beschreibenden Normalenvektor, den Kopplungsvektor \vec{J} , zu finden, der diesen Muster-Ergebnis-Satz reproduziert. Nach der Überlegung, dass das Vorzeichen von $S_\mu \vec{J} \cdot \vec{x}^{(\mu)}$ immer positiv sein muss, definieren wir ein

$$E_\mu := S_\mu \vec{J} \cdot \vec{x}^{(\mu)} \quad (1)$$

und stellen folgende Bedingung dafür auf:

$$E_\mu > 0, \quad \mu = 1, \dots, P$$

Es ist eine Annahme, dass der Vektor \vec{J} , der diese Bedingung für alle Mustervektoren $\vec{x}^{(\mu)}$ erfüllen soll, durch eine Mittelung über alle Muster $\vec{x}^{(\nu)}$, gewichtet mit ihrem Ergebnis S_ν , berechenbar ist:

$$\vec{J} = \frac{1}{N} \sum_{\nu=1}^P S_\nu \vec{x}^{(\nu)} \quad (2)$$

Auf diese Weise bestimmte Kopplungsvektoren nennt man Hebbsche Kopplungen.

Die Annahme soll im Folgenden überprüft werden:

setzt man (2) in (1) ein und nimmt dabei an, dass alle auf N normierten Mustervektoren $\vec{x}^{(\mu)}$ zufällige, voneinander unabhängige Größen sind, welche isotrop im N -dimensionalen Raum verteilt sind, erhält man einen Ausdruck für E_μ mit einer Summe über unabhängige und gleichverteilte Zufallsgrößen Z_ν :

$$E_\mu = 1 + \frac{1}{N} \sum_{\nu=1}^{P-1} Z_\nu$$

Der Grenzwert der Summe ist dann normalverteilt um den Mittelwert 0 . $E_\mu = 1$ ist also am wahrscheinlichsten. Betrachtet man die Varianz dieser Verteilung, stellt man fest, dass der Schwanz der Gaußglocke nicht vernachlässigbar in den negativen E_μ -Bereich hineinragt. $E_\mu < 0$ bedeutet die falsche Klassifikation eines Musters, sodass Hebbische Kopplungen keine garantiert richtige Reproduktion des gegebenen Muster-Ergebnis-Satzes liefern.

Der Perzeptron-Lernalgorithmus von Rosenblatt

1958 präsentierte Frank Rosenblatt seinen iterativen Algorithmus, der für jede linear separable Dichotomie einen Kopplungsvektor \vec{J} finden kann, mit dem keine falschen Klassifikationen möglich sind. Es lässt sich also ein Perzeptron finden, das das gegebene Problem garantiert richtig löst / alle Muster korrekt klassifiziert.

Bei einem Startwert (z.B. dem Nullvektor) beginnend, prüft der Algorithmus in jedem Iterationsschritt, ob der nächste Mustervektor $\vec{x}^{(\mu)}$ in der Liste, welche sequentiell durchlaufen wird, vom Kopplungsvektor richtig klassifiziert wird. Ist dies der Fall, bleibt der aktuelle Kopplungsvektor $\vec{J}(n)$ erhalten, ansonsten wird er folgendermaßen verändert:

$$\vec{J}(n+1) = \vec{J}(n) + \frac{1}{N} S_\mu \vec{x}^{(\mu)}$$

Ein solcher Update-Schritt des Kopplungsvektors ist ein Lernschritt des Perzeptrons. Die Liste an Mustern wird eventuell mehrfach durchlaufen, bis jeder Mustervektor richtig klassifiziert wird.

Man kann beweisen, dass die Anzahl k dieser Lernschritte mit Update beschränkt ist (was Rosenblatt auch getan hat):

$$k \leq \frac{N}{\kappa(\vec{J})^2} \quad \text{mit} \quad \kappa_\mu = \frac{S_\mu \vec{J} \vec{x}^{(\mu)}}{\|\vec{J}\|} \quad \text{und} \quad \kappa(\vec{J}) = \text{Min} \kappa_\mu$$

κ_μ ist die Projektion eines Mustervektors $\vec{x}^{(\mu)}$ auf einen korrekten, normierten Kopplungsvektor \vec{J} , den Normalenvektor zur Trennebene. Es handelt sich bei κ_μ also um den Abstand von $\vec{x}^{(\mu)}$ zur Hyperebene. $\kappa(\vec{J})$ ist somit der kleinste unter diesen. Die benötigte Zahl an Schritten, bis der Algorithmus ein korrektes \vec{J} gefunden hat, kann extrem groß sein, wenn $\kappa(\vec{J})$ sehr klein ist.

Das Perzeptron optimaler Stabilität

$\kappa(\vec{J})$ bezeichnet man auch als Stabilität des Perzeptrons. Wenn diese sehr klein ist, bewirken schon kleinste "Verdrehungen" der Trennebene, also Veränderungen der Kopplung, eine falsche Klassifikation bestimmter - nahe der Trennebene liegender - Muster.

Wenn es mehrere mögliche Einstellungen der Ebene gibt, sie also in einem bestimmten Bereich frei "drehbar" ist, ohne die gegebene Dichotomie zu verletzen, lässt sich unter Variation von \vec{J} der kleinste Muster-Ebene-Abstand $\kappa(\vec{J})$ maximieren. Nach dieser Optimierung hat das Perzeptron den Kopplungsvektor \vec{J}_{max} und ist das Perzeptron optimaler Stabilität für das gegebene Problem.

Optimierung des Iterationsalgorithmus

Der Algorithmus von Rosenblatt kann optimiert werden, indem die Update-Schritte auf $\vec{J}(n+1)$ immer nur für das momentan am schlechtesten klassifizierte $\vec{x}^{(\mu)}$ bei aktuellem $\vec{J}(n)$ durchgeführt werden. Dieses $\vec{x}^{(\mu)}$ zeichnet sich durch ein minimales κ_μ aus (mit negativem Vorzeichen).