Machine Learning Übungsblatt 5

Ramon Leiser Tobias Hahn

3. Januar 2017

1 Markov Decision Processes

2 Perzeptron

initial sind die Gewichte auf 0 gesetzt. Als Trainingsdaten für die Gewichtsanpassung per Gradientenabstieg werden die Belegungen der Variablen aus der Angabe benutzt. Da alle Daten angegeben werden sollen und die Trainingsdatenmenge äußerst gering ist, bietet es sich an die Berechnungen per Hand durchzuführen. Die zufällige Auswahl der Trainingsdaten wird per Hand gemacht.

2.1 Schritt 1

w_0	0
w_1	0
w_2	0

Zufällig ausgewähltes Beispiel (aus den vier Trainingsbeispielen) ist Beispiel 1. Input ist also 0 und 0, sowie Output 0. Zuerst muss die Aktivierung des Perzeptrons berechnet werden. Dazu brauchen wir den Wert y, der sich aus den Gewichten und den Eingabewerten ergibt, sowie den sich daraus ergebenden Wert der Aktivierungsfunktion.

$$y = w_0 * I_0 + w_1 * I_1 - w_2 = 0 * 0 + 0 * 0 - 0 = 0$$
$$\sigma(y) = \frac{1}{1 + e^{-y}} = 0.5$$

Nachdem nun die Aktivierung des Perzeptrons feststeht muss der Fehler berechnet werden. Hierzu verwenden wir den quadrierten Fehler, also das Quadrat des Unterschieds zwischen der erwarteten Ausgabe und der tatsächlichen.

$$Fehler = (o - O)^2 = (0.5 - 0)^2 = 0.25$$

Um nun die Gewichte anzupassen müssen wir uns zuerst den Gradienten ausrechnen. Der Gradient baut auf auf dem Fehler der Ausgabe sowie den Ableitungen der Aktivierungsfunktion. Außerdem baut er auf auf dem Input. Die Formeln für den Gradienten sind:

$$\frac{\delta E}{\delta w_0} = \sigma(y) * (1 - \sigma(y)) * Fehler * I_0 = 0.5 * 0.5 * 0.5 * 0 = 0$$
$$\frac{\delta E}{\delta w_1} = \sigma(y) * (1 - \sigma(y)) * Fehler * I_1 = 0.5 * 0.5 * 0.5 * 0 = 0$$
$$\frac{\delta E}{\delta w_2} = \sigma(y) * (1 - \sigma(y)) * Fehler * -1 = 0.5 * 0.5 * 0.5 * -1 = -0.125$$

Da der Gradient bei den ersten zwei Gewichten 0 ist nützt auch eine Anpassung nichts - nur beim letzten Gewicht, dem Biasgewicht, nehmen wir eine Anpassung vor. Hier lautet die Formel für das neue Gewicht:

$$w_2 = w_2 - \alpha * \frac{\delta E}{\delta w_2} = 0 - 10 * -0.125 = 0 + 1.25 = 1.25$$

2.2 Schritt 2

w_0	0
w_1	0
w_2	1.25

Zufällig ausgewähltes Beispiel (aus den vier Trainingsbeispielen) ist Beispiel 3. Input ist also 0 und 1, sowie Output 1. Zuerst muss die Aktivierung des Perzeptrons berechnet werden. Dazu brauchen wir den Wert y, der sich aus den Gewichten und den Eingabewerten ergibt, sowie den sich daraus ergebenden Wert der Aktivierungsfunktion.

$$y = w_0 * I_0 + w_1 * I_1 - w_2 = 0 * 0 + 0 * 1 - 1.25 = -1.25$$
$$\sigma(y) = \frac{1}{1 + e^{-y}} = 0.2227$$

Nachdem nun die Aktivierung des Perzeptrons feststeht muss der Fehler berechnet werden. Hierzu verwenden wir den quadrierten Fehler, also das Quadrat des Unterschieds zwischen der erwarteten Ausgabe und der tatsächlichen.

$$Fehler = (o - O)^2 = (0.2227 - 1)^2 = -0.7773^2 = 0.6$$

Um nun die Gewichte anzupassen müssen wir uns zuerst den Gradienten ausrechnen. Der Gradient baut auf auf dem Fehler der Ausgabe sowie den Ableitungen der Aktivierungsfunktion. Außerdem baut er auf auf dem Input. Die Formeln für den Gradienten sind:

$$\frac{\delta E}{\delta w_0} = \sigma(y) * (1 - \sigma(y)) * Fehler * I_0 = 0.2227 * 0.7773 * -0.7773 * 0 = 0$$
$$\frac{\delta E}{\delta w_1} = \sigma(y) * (1 - \sigma(y)) * Fehler * I_1 = 0.2227 * 0.7773 * -0.7773 * 1 = -0.1346$$
$$\frac{\delta E}{\delta w_2} = \sigma(y) * (1 - \sigma(y)) * Fehler * -1 = 0.2227 * 0.7773 * -0.7773 * -1 = 0.1346$$

Da der Gradient beim ersten Gewicht 0 ist nützt auch eine Anpassung nichts - nur bei den letzten zwei Gewichten nehmen wir eine Anpassung vor. Hier lautet die Formel für das neue Gewicht:

$$w_1 = w_1 - \alpha * \frac{\delta E}{\delta w_1} = 0 - 10 * -0.1346 = 0 + 1.346 = 1.346$$
$$w_2 = w_2 - \alpha * \frac{\delta E}{\delta w_2} = 1.25 - 10 * 0.1346 = 1.25 - 1.346 = -0.096$$

2.3 Schritt 3

w_0	0
w_1	1.346
w_2	-0.096

Zufällig ausgewähltes Beispiel (aus den vier Trainingsbeispielen) ist Beispiel 2. Input ist also 1 und 0, sowie Output 1. Zuerst muss die Aktivierung des Perzeptrons berechnet werden. Dazu brauchen wir den Wert y, der sich aus den Gewichten und den Eingabewerten ergibt, sowie den sich daraus ergebenden Wert der Aktivierungsfunktion.

$$y = w_0 * I_0 + w_1 * I_1 - w_2 = 1 * 0 + 0 * 1.346 + 0.096 = 0.096$$
$$\sigma(y) = \frac{1}{1 + e^{-y}} = 0.524$$

Nachdem nun die Aktivierung des Perzeptrons feststeht muss der Fehler berechnet werden. Hierzu verwenden wir den quadrierten Fehler, also das Quadrat des Unterschieds zwischen der erwarteten Ausgabe und der tatsächlichen.

$$Fehler = (o - O)^2 = (0.524 - 1)^2 = -0.476^2 = 0.227$$

Um nun die Gewichte anzupassen müssen wir uns zuerst den Gradienten ausrechnen. Der Gradient baut auf auf dem Fehler der Ausgabe sowie den Ableitungen der Aktivierungsfunktion. Außerdem baut er auf auf dem Input. Die Formeln für den Gradienten sind:

$$\frac{\delta E}{\delta w_0} = \sigma(y) * (1 - \sigma(y)) * Fehler * I_0 = 0.227 * 0.476 * -0.476 * 1 = -0.05$$
$$\frac{\delta E}{\delta w_1} = \sigma(y) * (1 - \sigma(y)) * Fehler * I_1 = 0.227 * 0.476 * -0.476 * 0 = 0$$
$$\frac{\delta E}{\delta w_2} = \sigma(y) * (1 - \sigma(y)) * Fehler * -1 = 0.227 * 0.476 * -0.476 * -1 = 0.05$$

Da der Gradient beim zweiten Gewicht 0 ist nützt auch eine Anpassung nichts - nur bei den anderen zwei Gewichten nehmen wir eine Anpassung vor. Hier lautet die Formel für das neue Gewicht:

$$w_0 = w_0 - \alpha * \frac{\delta E}{\delta w_0} = 0 - 10 * -0.05 = 0 + 0.5 = 0.5$$
$$w_2 = w_2 - \alpha * \frac{\delta E}{\delta w_2} = -0.096 - 10 * 0.05 = -0.096 - 0.5 = -0.596$$

2.4 Schritt 3

w_0	0.5
w_1	1.346
w_2	-0.596

Zufällig ausgewähltes Beispiel (aus den vier Trainingsbeispielen) ist Beispiel 4. Input ist also 1 und 1, sowie Output 1. Zuerst muss die Aktivierung des Perzeptrons berechnet werden. Dazu brauchen wir den Wert y, der sich aus den Gewichten und den Eingabewerten ergibt, sowie den sich daraus ergebenden Wert der Aktivierungsfunktion.

$$y = w_0 * I_0 + w_1 * I_1 - w_2 = 0.5 * 1 + 1.346 * 1 + 0.596 = 2.442$$
$$\sigma(y) = \frac{1}{1 + e^{-y}} = 0.92$$

Nachdem nun die Aktivierung des Perzeptrons feststeht muss der Fehler berechnet werden. Hierzu verwenden wir den quadrierten Fehler, also das Quadrat des Unterschieds zwischen der erwarteten Ausgabe und der tatsächlichen.

$$Fehler = (o - O)^2 = (0.92 - 1)^2 = -0.08^2 = 0.0064$$

Damit ist der quadrierte Fehler unter 0.1 gesunken, was in der Angabe als Abbruchbedingung eingeführt wurde. Damit sind die Gewichte, die wir herausgefunden haben, auch die tatsächlichen, und wir können mit den Berechnungen aufhören.

2.5 Beispiele

Laut Angabe sollen noch zwei Beispieleingaben durchgerechnet werden, und die Ergebnisse interpretiert. Dies geschieht hier.

2.5.1 0.5

Beide Eingaben sind 0.5. Damit ergibt sich folgende Berechnung:

$$y = w_0 * I_0 + w_1 * I_1 - w_2 = 0.5 * 0.5 + 1.346 * 0.5 + 0.596 = 1.519$$
$$\sigma(y) = \frac{1}{1 + e^{-y}} = 0.82$$

Hier entspricht die Berechnete Wahrscheinlichkeit ungefähr dem Wert der erwartet wird. Dies begründet sich so: Bei einer Wahrscheinlichkeit von 0.5 ist die Wahrscheinlichkeit für 1 oder 0 gleich wahrscheinlich, d.h. der Eingabewert entspricht dem Münzwurf einer perfekten Münze. Hier können nun vier verschiedene Ausgaben vorkommen, so wie in der Tabelle in der Angabe angegeben. Bei drei dieser Ausgaben bekommen wir einen Wert 1, bei einer 0, d.h. der Erwartungswert für die Ausgabe ist 0.75, was recht nahe an unserem Ergebnis ist.

2.5.2 0.1

Beide Eingaben sind 0.1. Damit ergibt sich folgende Berechnung:

$$y = w_0 * I_0 + w_1 * I_1 - w_2 = 0.5 * 0.1 + 1.346 * 0.1 + 0.596 = 0.78$$
$$\sigma(y) = \frac{1}{1 + e^{-y}} = 0.69$$

Diese Berechnung deckt sich nicht mit dem Wert den wir erwarten. Die Wahrscheinlichkeit dass einer der Eingaben 0 ist beträgt 0.9, d.h. dass beide Eingaben 0 sind 0.9 * 0.9 oder 0.81. In allen anderen Fällen ist die Ausgabe 1, d.h. unsere Ausgabe sollte 1 - 0.81 oder 0.19 sein. Davon ist unser Wert um 0.5 entfernt, was eine große Abweichung darstellt. Dies liegt daran, dass das Perzeptron nicht genug trainiert wurde, da immer nur ein Beispiel gewählt wurde, und es zu zufälligen sehr guten Ergebnissen bei einem Beispiel kommen kann. Besser wäre es gewesen alle Beispiele auf einmal zu lernen, dann hätte die Summe der Fehler etwas mehr ausgesagt, dan wäre aber mit Hand rechnen sehr kompliziert gewesen.

2.6 Lernrate

Die Lernrate hat einen äußerst großen Einfluss auf den Gradientenabstieg. Um diesen Einfluss zu verstehen muss man zuerst einmal verstehen was ein Gradient ist. Ein Gradient ist im Grunde nichts anderes als ein Vektor, der in eine Richtung zeigt. Diese Richtung ist die Richtung, in der man die jeweiligen Elemente des Vektors verändern muss, um die Funktion für die der Gradient berechnet wurde zu maximieren (deswegen nimmt man bei der Gewichtsanpassung immer den negativen Gradienten). Die Lernrate gibt dann an, wie weit man in diese Richtung geht. Schon hier kann man erkennen dass die Lernrate eine große Rolle spielt, jedoch muss man sich noch verdeutlichen welche genauen Auswirkungen das hat.

Nehmen wir dafür einmal an wir haben eine zu große Lernrate gewählt. In diesem Fall machen wir einen so weiten Schritt in die Richtung des Minimums, dass wir das Minimum überschreiten und auf der anderen Seite den Funktionswert wiederum vergrößern statt verkleinern. Das kann die Auswirkung haben dass wir einen verlängerten Abstieg haben, aber immer noch absteigen - schreiten wir jedoch so weit über das Minimum hinaus, dass der Fehlerwert tatsächlich größer wird als beim Ausgangspunkt, so kann es passieren dass der Fehler sich immer weiter aufschaukelt (bei konkaven Funktionen ist dies sogar garantiert).

Dagegen ist die Lernrate zu klein gewählt: Hier wird sich zwar der Fehler nie automatisch vergrößern, jedoch machen wir einen so kleinen Schritt in die Richtung des Minimums dass unzählige Epochen notwendig sind um eine Verbesserung zu spüren. Dies führt dazu dass unser Algorithmus endlos lange rechnen muss bis er eine aktzeptable Lösung gefunden hat.