Universidad Simón Bolívar
Departamento de Cómputo Científico

Tarea I: Calentamiento

Georvic Tur - 12-11402

25/04/2017

1. La función logística se define como

$$\varphi(v) = \frac{1}{1 + \exp(-v)} \tag{1}$$

con rango entre 0 y 1. Muestre que la derivada satisface la siguiente relación:

$$\frac{\mathrm{d}\varphi}{\mathrm{d}v} = \varphi(v)[1 - \varphi(v)] \tag{2}$$

¿Cuál es el valor de la derivada en el origen?

Solución:

Derivemos paso por paso la función φ. Si aplicamos el teorema de la derivada de una fracción, tenemos que:

$$\begin{aligned}
\frac{\mathrm{d}\varphi}{\mathrm{d}v} &= \frac{-\frac{\mathrm{d}(1+exp(-v))}{\mathrm{d}v}}{(1+exp(-v))^2} && \text{(derivada de una fracción)}\\
&= \frac{exp(-v)}{(1+exp(-v))^2} && \text{(regla de la cadena)}\\
&= (\frac{1}{1+exp(-v)}) * (\frac{exp(-v)}{1+exp(-v)}) && \text{(reordenando)}\\
&= \varphi(v) * (\frac{exp(-v)}{1+exp(-v)}) && \text{(Sustituyendo la ecuación 1)}\\
&= \varphi(v) * (\frac{exp(-v) + (1-1)}{1+exp(-v)}) && \text{(Elemento neutro de la adición)}\\
&= \varphi(v) * (1 - \frac{1}{1+exp(-v)}) && \text{(Simplificando)}\\
&= \varphi(v) * (1 - \varphi(v)) && \text{(Sustituyendo la ecuación 1)}
\end{aligned}$$

Ahora calculemos $\frac{\mathrm{d}\varphi}{\mathrm{d}v}|_{v=0}$:

$$
\begin{aligned}
\frac{d\varphi}{dv}|_{v=0} &= \varphi(0) * (1-\varphi(0)) && \text{(derivada de la logística)}\\
&= \frac{1}{2} * (1-\frac{1}{2}) && \text{(Valor de la logística en el origen)}\\
&= \frac{1}{4}
\end{aligned}
$$

2. En la pregunta anterior vimos que la derivada de la función logística (φ) se puede expresar como función de sí misma. Encuentre un resultado similar para la función de activación de la tangente hiperbólica definida como:

$$
\tanh(v) = \frac{\exp(v) - \exp(-v)}{\exp(v) + \exp(-v)} \tag{3}
$$

Solución:

Calculemos su derivada:

$$
\begin{aligned}
\tanh(v) &= \frac{(\exp(v)+\exp(-v))(\exp(v)+\exp(-v)) - (\exp(v)-\exp(-v))(\exp(v)-\exp(-v))}{(\exp(v)+\exp(-v))^2}\\
& && \text{(derivada de una fracción)}\\
&= \frac{(\exp(v)+\exp(-v))^2 - (\exp(v)-\exp(-v))^2}{(\exp(v)+\exp(-v))^2} && \text{(Simplificando)}\\
&= 1 - \Big(\frac{\exp(v)-\exp(-v)}{\exp(v)+\exp(-v)}\Big)^2 && \text{(Simplificando)}\\
&= 1 - \tanh(v) && \text{(Ecuación 3)}
\end{aligned}
$$

Ahora calculemos $\frac{d\tanh}{dv}|_{v=0}$:

$$
\begin{aligned}
\frac{d\tanh}{dv}|_{v=0} &= 1 - \tanh(0) && \text{(Derivada de tanh)}\\
&= 1 && \text{(Tangente hiperbólica es cero en el origen)}
\end{aligned}
$$

3. Considere un grupo de personas cuyas opiniones colectivas en un tópico de interés se define como la suma pesada de las opiniones individuales. Suponga que si en el transcurir del tiempo la opinión de un miembro del grupo tiende a (concuerda con) la opinión colectiva del grupo, la opinión de ese miembro recibe un mayor peso. Si por el contrario, la opinión de ese miembro

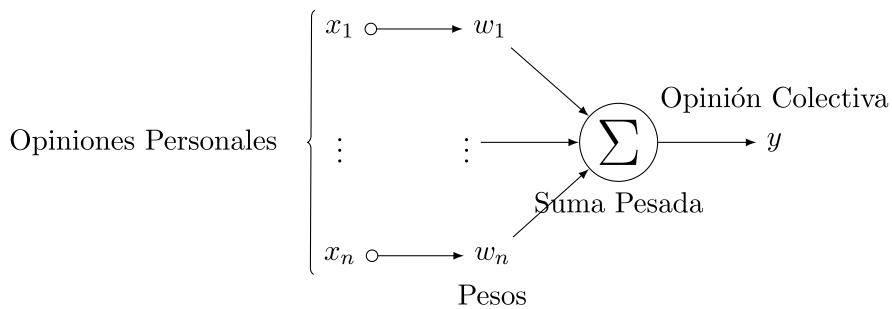
consistentemente está en desacuerdo con la opinión colectiva, la opinión de dicho miembro se le da menos peso. Esto tiene el efecto de producir consenso en el grupo. Discuta la analogía de esta situación con alguno de los postulados de aprendizaje.

Solución:

Esto es similar al postulado de Hebb parafraseado por Haykin en su libro [1]:

1. Si dos neuronas se activan simultáneamente, la fuerza de su sinapsis se incrementa
2. Si dos neuronas se activan de manera asíncrona, entonces la fuerza de su sinapsis disminuye

En este caso tendríamos (Haykin, 3rd ed, p. 370)



Ese sería también uno de los principios de la auto-organización (Haykin, 3rd ed, p. 368)

4. Para una red neuronal de la forma

$$y_k = \sigma(\sum_{j=1}^{K} w_{kj}^{(2)} \sigma(\sum_{i=1}^{M} w_{ji}^{(1)} x_i)) \tag{4}$$

en la cual la función de activación σ está dada por función logística, muestre que existe una red equivalente salvo por transformaciones lineales de los parámetros del modelo, que calcula exactamente igual pero con función de activación en la capa oculta dada por la tangente hiperbólica. Ayuda: Primero encuentre una relación entre la logística y la tangente hiperbólica.

Solución:

Supongamos que la función logística depende de tanh por las formas de sus gráficas: $\varphi(v) = a + b * \tanh(c * v + d)$. Si aplicamos límites, tenemos:

$$\begin{aligned}
\lim_{v\to\infty} \varphi(v) &= 1 && \text{(Límite de la logística al tender al infinito)}\\
a + b * (\lim_{v\to\infty} \tanh(c * v + d)) &= 1 && \text{(Sustituyendo la suposición y aplicando límite)}\\
a + b &= 1 && \text{(Límite de tanh al tener al infinito)}
\end{aligned}$$

Asimismo:

$$\begin{aligned}
\lim_{v\to-\infty} \varphi(v) &= 0 && \text{(Límite de la logística al tender al infinito negativo)}\\
a + b * (\lim_{v\to-\infty} \tanh(c * v + d)) &= 0 && \text{(Sustituyendo la suposición y aplicando límite)}\\
a - b &= 0 && \text{(Límite de tanh al tender al infinito negativo)}
\end{aligned}$$

Con esas dos ecuaciones tenemos $a = b = \frac{1}{2}$. Si evaluamos $\varphi(0) = \frac{1}{2}$, tenemos:

$$\begin{aligned}
\varphi(0) &= \frac{1}{2} && \text{(Valor de la logística)}\\
a + b * \tanh(0 + d) &= \frac{1}{2} && \text{(Suposición)}\\
\frac{1}{2} + \frac{1}{2} * \tanh(0 + d) &= \frac{1}{2} && \text{(Valores de a y b)}\\
\tanh(d) &= 0 && \text{(Simplificando)}\\
&\Longrightarrow \\
d &= 0 && \text{(Valor de tanh en el origen)}
\end{aligned}$$

Si aplicamos la derivada evaluada en cero, tenemos:

$$\begin{aligned}
\frac{\mathrm{d}\varphi}{\mathrm{d}v}\Big|_{v=0} &= \frac{1}{4} && \text{(Por la pregunta 1)}\\
b * \left(\frac{\mathrm{d}(\tanh(c * v + d))}{\mathrm{d}v}\Big|_{v=0}\right) &= \frac{1}{4} && \text{(Por suposición y derivada de una constante)}\\
b * c * \Big(1 - tanh(c * 0 + d)\Big)^2 &= \frac{1}{4} && \text{(Por pregunta 2 y regla de la cadena)}\\
\frac{1}{2} * c * \Big(1 - tanh(0)\Big)^2 &= \frac{1}{4} && \text{(Valores de a, b y d)}\\
\frac{1}{2} * c &= \frac{1}{4} && \text{(Valor de tanh en el origen)}\\
&\Longrightarrow \\
c &= \frac{1}{2}
\end{aligned}$$

Por tanto, tenemos que $\varphi(v) = \frac{1}{2} + \frac{1}{2}\tanh\left(\frac{1}{2}v\right)$. Si sustituimos a v por $2v$, podemos obtener:

$$\tanh(v) = 2\varphi(2v) + 1 \tag{5}$$

Con ayuda de la ecuación 5, podemos hacer lo siguiente:

$$\begin{aligned}
y_k &= \sigma(\sum_{j=1}^{K} w_{kj}^{(2)} \tanh\left(\sum_{i=1}^{M} w_{ji}^{(1)} x_i\right)) && \text{(Enunciado)} \\
&= \sigma(\sum_{j=1}^{K} w_{kj}^{(2)} \Big(2\varphi(2\sum_{i=1}^{M} w_{ji}^{(1)}) + 1\Big)) && \text{(Ecuación 5)} \\
&= \sigma(\sum_{j=1}^{K} (2w_{kj}^{(2)})\varphi(\sum_{i=1}^{M}(2w_{ji}^{(1)})) + \sum_{j=1}^{K} w_{kj}^{(2)}) && \text{(Distribuyendo)}
\end{aligned}$$

Donde $2w_{ji}^{(1)}$ y $2w_{kj}^{(2)}$ son transformaciones lineales de los parámetros de la red neuronal y $b = \sum_{j=1}^{K} w_{kj}^{(2)}$ puede considerarse un bias. Por tanto, hemos mostrado que una red de dos capas como la del enunciado es equivalente a otra que use en la capa oculta a tanh.

Referencias

[1] Simon Haykin, *Neural networks and learning machines, 3rd ed.* Upper Saddle River: Pearson Education, 2009. p, 369.