Autoencoder

Se tiene una función que trata de aprender una aproximación a la función de identidad, es decir una red neuronal que genera un vector $\widehat{\mathbb{X}}_i$ que sea similar al vector de entrada $\mathbb{X}_i$.

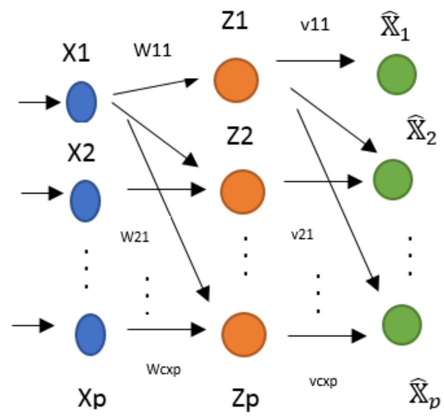

Figura 1. Red neuronal de un autoencoder

$$sea\ \ el\ verctor\ de\ dato\ de\ entrada\ , \mathbb{X}_i = \begin{pmatrix} \mathbb{X}_{1i} \\ \mathbb{X}_{2i} \\ \vdots \\ \mathbb{X}_{pi} \end{pmatrix}$$

$$sea\ el\ vertor\ de\ salida\ de\ la\ capa\ oculta, \mathbb{z}_i = \begin{pmatrix} \mathbb{z}_{1i} \\ \mathbb{z}_{2i} \\ \vdots \\ \mathbb{z}_{pi} \end{pmatrix}$$

$$\text{Se tiene entonces la matriz de pesos de los datos de entrada } W_{cxp} = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1p} \\ w_{21} & w_{22} & \cdots & w_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ w_{c1} & w_{c2} & \cdots & w_{cp} \end{pmatrix}$$

$$sea\ f\ una\ funcion\ de\ activacion\ lineal, \mathbb{z}_i = f(W\mathbb{X}_i)$$

$$por\ ello\ \ se\ tiene\ \ \ \mathbb{z}_i = W\mathbb{X}_i$$

Para el vector de datos estimados $\widehat{\mathbb{X}}_i$,se tiene v matriz de pesos de la salida de la capa oculta y g una función de activación lineal:

$$\widehat{\mathbb{X}}_i = g(vf(W\mathbb{X}_i)) \qquad g \text{ y } f \text{ al ser lineales}$$

Entonces: $\widehat{\mathbb{X}}_i = VW\mathbb{X}_i$ (1)

Definimos una función de costo: $J = \frac{1}{N}\sum_{i=0}^{N-1}\left|\widehat{\mathbb{X}}_i - \mathbb{X}_i\right|^2$ y remplazamos $\widehat{\mathbb{X}}_i$ de (1)

$$J = \frac{1}{N}\sum_{i=0}^{N-1}|VW\mathbb{X}_i - \mathbb{X}_i|^2$$

Se optimiza la función de costo obtenida:

$$\nabla J(VW) = \frac{1}{N}2\sum_{i=0}^{N-1}|VW\mathbb{X}_i - \mathbb{X}_i|^2\,\mathbb{X}_i = 0$$

$$= \left(\frac{1}{N}2\sum_{i=0}^{N-1}|\mathbb{X}_i|^2\right)|VW - I| = 0$$

Para que la expresión anterior se cumpla, se debe encontrar los vectores y valores propios de la siguiente expresión:

$$|VW - I| = 0$$

$$|\lambda VW - I\lambda| = 0 \;\; (2)$$

Análisis de componentes principales (PCA)

En el análisis de componentes principales se busca reducir la dimensión de un conjunto $\mathbb{X}_{pi}$ de variables, conservando la mayor cantidad de información que sea posible:

$$\mathbb{X}_{pi} \rightarrow \mathbb{Z}_{ci} \quad \text{donde } \; c < p$$

Sea entonces la A matriz de componentes principales

$$\mathbb{z}_i = A\mathbb{X}_i$$

Y de este modo se tiene el vector de componentes principales $\mathbb{z}_i$:

$$\mathbb{z}_i = \begin{pmatrix} \hat{a}_1\mathbb{X}_i \\ \hat{a}_2\mathbb{X}_i \\ \vdots \\ \hat{a}_c\mathbb{X}_i \end{pmatrix} = \begin{pmatrix} \mathbb{z}_1 \\ \mathbb{z}_2 \\ \vdots \\ \mathbb{z}_c \end{pmatrix}$$

Tomamos la componente de mayor variabilidad

$$\mathbb{z}_1 = \hat{a}_1\mathbb{X}_i$$

Encontramos su varianza:

$$var(\mathbb{z}_1) = var(\hat{a}_1\mathbb{X}_i)$$

$$var(\mathbb{z}_1) = \hat{a}_1{}^t\Sigma\hat{a}_1$$

Donde $\hat{a}_1$ ortogonal y Σ la matriz de varianzas-covarianzas

$$\hat{a}_1{}^t\hat{a}_1 = 1$$

Para analizar la varianza del vector aplicamos los multiplicadores de LaGrange:

$$L(\hat{a}) = a^t\,\Sigma a - \lambda(a^t a - 1), a^t a = |a|^2$$

Y de este modo encontrar la máxima varianza derivamos los multiplicadores de LaGrange e igualamos 0.

$$\frac{\partial L(\hat{a})}{\partial \hat{a}} = 2\Sigma a - 2\lambda a = 0$$

$$= 2(\Sigma - \lambda I)a = 0$$

De este modo la expresión anterior se cumple, encontrando los vectores y valores propios de la siguiente expresión

$$\Sigma - \lambda I = 0 \text{ (3)}$$

Se puede observar que para el autoencoder aprenda una aproximación a la función de identidad, se debe encontrar los pesos de la expresión (2) y de forma similar para el PCA exista una reducción de dimensión de los datos de entrada, se calculan los valores y vectores de (3), los cuales son análogos con los pesos encontrados en (2).