Formulario del examen de mitad de curso

Deep Learning para Business Analytics · Comillas ICADE · Días 01–12

Neurona y red

Neurona

$$y = \sigma\left(\sum_{i=1}^{d} w_i x_i + b\right) = \sigma(w^\top x + b)$$

Perceptrón multicapa

$$f_\theta(x) = h_L \circ \cdots \circ h_1(x), \qquad h_\ell(z) = \sigma_\ell(W_\ell z + b_\ell)$$

Sigmoide

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

ReLU

$$\mathrm{ReLU}(z) = \max(0, z)$$

Normalización

BatchNorm

$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} a_i, \qquad \sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m}(a_i - \mu_B)^2$$

BatchNorm: salida

$$\hat{a} = \frac{a - \mu_B}{\sigma_B}, \qquad y = \gamma \hat{a} + \beta$$

Convolución

Convolución 2D

$$Y_{i,j,o} = \sum_{u=0}^{k-1}\sum_{v=0}^{k-1}\sum_{c=1}^{C_{in}} X_{i+u,\, j+v,\, c}\, K_{u,v,c,o} + b_o$$

Dimensión espacial de salida

$$H' = \left\lfloor \frac{H + 2p - k}{s} \right\rfloor + 1$$

Max-pooling

$$Y_{i,j,c} = \max_{0 \le u,v < k_p} X_{i\cdot s+u,\, j\cdot s+v,\, c}$$

Average pooling

$$Y_{i,j,c} = \frac{1}{k_p^2}\sum_{u=0}^{k_p-1}\sum_{v=0}^{k_p-1} X_{i\cdot s+u,\, j\cdot s+v,\, c}$$

Modelos secuenciales

RNN

$$h_t = \sigma\left(W_h h_{t-1} + W_x x_t + b\right)$$

State space model

$$h_t = \bar{A} h_{t-1} + \bar{B} x_t, \qquad y_t = C h_t$$

Atención

Atención escalada

$$\alpha_j = \mathrm{softmax}\left(\frac{q^\top k_j}{\sqrt{d_k}}\right), \qquad \mathrm{Att} = \sum_j \alpha_j\, v_j$$

Embeddings

Similitud coseno

$$\cos(u, v) = \frac{u^\top v}{\|u\|_2\, \|v\|_2} = \frac{\sum_i u_i v_i}{\sqrt{\sum_i u_i^2}\,\sqrt{\sum_i v_i^2}}$$

Aritmética semántica

$$\phi(\text{rey}) - \phi(\text{hombre}) + \phi(\text{mujer}) \approx \phi(\text{reina})$$

Modelos de lenguaje

Temperatura

$$p_\tau(x_i) = \frac{\exp(\ell_i/\tau)}{\sum_j \exp(\ell_j/\tau)}$$

Difusión

Factores del schedule

$$\alpha_t = 1 - \beta_t, \qquad \bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$$

Marginal del proceso forward

$$q\left(x_t \,\middle|\, x_0\right) = \mathcal{N}\left(x_t;\ \sqrt{\bar{\alpha}_t}\, x_0,\ (1 - \bar{\alpha}_t) I\right)$$