Quiz 2-Precursor

CS-583: Deep Learning

Wyatt Blair

April 16, 2024

Problem 1. Feed-Forward Neural Network (FNN)

(a) Model

$f_l(X_l) = a_l(w_l^T \cdot X_l); \forall l \in$ hidden-layers

$l :=$ hidden layer index
$X_l :=$ input data which has been transformed by $(l-1)$ hidden layers
$a_l :=$ activation function for layer l
$w_l :=$ weights for hidden layer l

Problem 2. Convolutional Neural Network (CNN)

(a) Convolutional Layer

$G[m,n] = (f * h)[m,n] = \sum_j \sum_k h[j,k] f[m-j, n-k]$

$f :=$ input image
$h :=$ kernel
$m, n :=$ matrix indices
$G :=$ resulting feature map

Problem 3. Recurrent Neural Network

(a) Model

$h_t = f_W(h_{t-1}, x_t)$

$h_t :=$ new state
$f_W :=$ function with parameters W
$h_{t-1} :=$ previous state
$x_t :=$ input vector at time step t

(b) Loss Function

$\ell(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \log y_{ic}$

Problem 4. Generative Adversarial Network (GAN)

(a) Discriminator Weight Gradient

$\nabla_w D_w = \frac{1}{m} \sum_{i=1}^{m} [\log D_w(x_i) + \log(1 - D_w(G_\theta(z_i)))]$

w := weights of the discriminator
θ := weights of the generator
m := number of samples in mini-batch
D_w := discriminator using weights w. Output restricted to $[0, 1]$, indicating whether the discriminator believes the given image, x, is fake or real.
G_θ := generator using weights θ. Output is an image which is designed to trick the discriminator.
x := an image in the discriminator's input space
z := a feature vector in the generator's input space

(b) Generator Weight Gradient

$\nabla_\theta G_\theta = \frac{1}{m} \sum_{i=1}^{m} \log(1 - D_w(G_\theta(z_i)))$

Please refer to the Discriminator section above for a description of the variables.

Problem 5. Autoencoder

(a) Model

$E_\phi : \mathcal{X} \to \mathcal{Z}$
$z = E_\phi(x)$

$D_\theta : \mathcal{Z} \to \mathcal{X}$
$x' = D_\theta(z)$

E_ϕ := encoder, parameterized by ϕ
D_θ := decoder, parameterized by θ
x := input message
z := latent code
x' := decoded message

(b) Objective Function

$L(\theta, \phi) := \mathbb{E}_{x\ \mu_{\text{ref}}}[d(x, D_\theta(E_\phi(x)))]$

$$\min_{\theta,\phi} L(\theta, \phi); \text{where} L(\theta, \phi) = \frac{1}{N} \sum_{i=1}^{N} ||x_i - D_\theta(E_\phi(x_i))||_2^2$$

Problem 6. Transformer

(a) Model

$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V$

Q := Query matrix
K := Key matrix
V := value matrix
d_k := number of dimensions