گزارش پروژه شبکه های عصبی

بخش اول – لایه ها

* 1. لایه Fully Connected :

تکمیل تابع وزن دهی:

* اگر متد وزن دهی رندم باشد: به صورت زیر فراخوانی تابع رندم نامپای انجام شده و وزن های رندوم انتخاب میشوند، سپس با ضرب کردن آنها در 0.01 آنها را اسکیل میکنیم.
* if self.initialize\_method == "random":  
   # Initialize weights with random values using np.random.randn  
   return np.random.randn(self.output\_size, self.input\_size) \* 0.01
* اگر متد وزن دهی xavier باشد: مقداردهی اولیه xavier یک روش رایج برای مقداردهی اولیه وزن‌ها در یک لایه شبکه عصبی است که می‌تواند به شبکه کمک کند در طی فرآیند آموزش به سرعت به جواب مسئله برسد. این روش از یک توزیع گاوسی برای مقداردهی اولیه وزن‌ها استفاده می‌کند که واریانس آن به نسبت جذر معکوس تعداد ورودی‌ها به لایه جاری به اضافه تعداد خروجی‌های آن لایه است.

در این قطعه کد، متغیر self.input\_size تعداد نورون‌های لایه قبلی را نشان می‌دهد، در حالی که self.output\_size تعداد نورون‌های لایه جاری را نمایش می‌دهد.کد، مقدار واریانس مورد نیاز برای مقداردهی اولیه وزن‌ها را با استفاده از فرمول مربوط به روش مقداردهی اولیه زاویه‌گیری محاسبه می‌کند و سپس با استفاده از تابع np.random.randn در NumPy، یک ماتریس تصادفی با ابعاد (self.output\_size, self.input\_size) تولید کرده و آن را در ضریب محاسبه شده برای واریانس، ضرب می‌کند. در نهایت، این ماتریس به عنوان وزن‌های مقداردهی اولیه لایه برگردانده می‌شود

elif self.initialize\_method == "xavier":  
 # Initialize weights using Xavier initialization  
 xavier\_stddev = np.sqrt(2 / (self.input\_size + self.output\_size))  
 return np.random.randn(self.output\_size, self.input\_size) \* xavier\_stddev

* اگر متد وزن دهی he باشد: مقداردهی اولیه he نیز یک روش رایج برای مقداردهی اولیه وزن‌های یک لایه شبکه عصبی است که به شبکه کمک می‌کند در طی فرآیند آموزش به سرعت به جواب مسئله برسد. در این روش، مقدار واریانس اولیه وزن‌ها برابر با دو برابر نسبت جذر معکوس تعداد ورودی‌ها به لایه جاری است. در این قطعه کد، متغیر self.input\_size تعداد نورون‌های لایه قبلی را نشان می‌دهد. ابتدا با استفاده از فرمول مربوط به روش مقداردهی اولیه هی، واریانس اولیه وزن‌ها محاسبه می‌شود و سپس با استفاده از تابع np.random.randn در NumPy، یک ماتریس تصادفی با ابعاد (self.output\_size, self.input\_size) تولید می‌شود و در ضریب محاسبه شده برای واریانس ضرب می‌شود. در نهایت، این ماتریس به عنوان وزن‌های مقداردهی اولیه لایه برگردانده می‌شود.
* elif self.initialize\_method == "he":  
   # Initialize weights using He initialization  
   he\_stddev = np.sqrt(2 / self.input\_size)  
   return np.random.randn(self.output\_size, self.input\_size) \* he\_stddev

تکمیل تابع تعیین بایاس:

def initialize\_bias(self):  
 # Initialize bias with zeros  
 return np.zeros((self.output\_size, 1))

در این تابع، با استفاده از تابع np.zeros() در پکیج NumPy، برداری به اندازه تعداد نورون‌های لایه جاری که با self.output\_size مشخص شده است، تولید می‌شود. اندازه دوم این بردار برابر یک است، زیرا هر بایاس یک مقدار تکی دارد که برای هر نورون اعمال می‌شود. سپس این بردار به عنوان بایاس مقداردهی اولیه لایه برگردانده می‌شود

تکمیل تابع Forward :

# NOTICE: BATCH\_SIZE is the first dimension of A\_prev  
self.input\_shape = A\_prev.shape  
A\_prev\_tmp = np.copy(A\_prev)

در این کد ماتریسی مشابه لایه نورون های ورودی ساخته میشود

# Check if A\_prev is output of convolutional layer  
if len(A\_prev\_tmp.shape) > 2:  
 batch\_size = A\_prev\_tmp.shape[0]  
 A\_prev\_tmp = A\_prev\_tmp.reshape(batch\_size, -1).T  
else:  
 batch\_size = A\_prev\_tmp.shape[0]  
  
self.reshaped\_shape = A\_prev\_tmp.shape

سپس چک میکنیم که این ورودی ها از یک لایه کانولوشنی آمده اند یا خیر. اگر ورودی بیش از دو بعد داشته باشد خروجی یک لایه کانولوشنی است. علت اهمیت بررسی این موضوع این است که در شبکه‌های عصبی، لایه‌های کانولوشنی و لایه‌های پرسپترون کامل (fully connected) به طور معمول به صورت متناوب در کنار هم قرار می‌گیرند. ورودی لایه‌های پرسپترون کامل (fully connected) باید به صورت یک بردار یک بعدی باشد، در حالی که ورودی لایه‌های کانولوشنی به صورت یک آرایه چند بعدی است. بنابراین، اگر ورودی لایه فعلی یک آرایه چند بعدی باشد (مانند خروجی لایه کانولوشنی)، لازم است آن را به یک بردار یک بعدی تبدیل کرد تا به لایه پرسپترون کامل (fully connected) داده شود. به همین دلیل، در این قسمت از کد، با بررسی تعداد بعد‌های ورودی، از اینکه ورودی لایه کانولوشنی باشد یا خیر، اطمینان حاصل شده و اگر ورودی یک آرایه چند بعدی باشد، آن را به یک بردار یک بعدی تبدیل می‌کند.

# Forward part  
W, b = self.weights, self.biases  
Z = np.dot(W, A\_prev\_tmp) + b  
return Z

در این مرحله مرحله فوروارد شبکه fully connected پیاده سازی شده است، همانطور که مشاهده میشود z حاصل دات شدن وزن ها و نورون های ورودی و اعمال بایاس میباشد.

تکمیل تابع backward :

def backward(self, dZ, A\_prev):  
 *"""  
 Backward pass for fully connected layer.  
 args:  
 dZ: derivative of the cost with respect to the output of the current layer  
 A\_prev: activations from previous layer (or input data)  
 returns:  
 dA\_prev: derivative of the cost with respect to the activation of the previous layer  
 grads: list of gradients for the weights and bias  
 """* A\_prev\_tmp = np.copy(A\_prev)  
  
 # Check if A\_prev is output of convolutional layer  
 if len(A\_prev\_tmp.shape) > 2:  
 batch\_size = A\_prev\_tmp.shape[0]  
 A\_prev\_tmp = A\_prev\_tmp.reshape(batch\_size, -1).T  
 else:  
 batch\_size = A\_prev\_tmp.shape[0]  
  
 # Backward part  
 W, b = self.weights, self.biases  
 dW = np.dot(dZ, A\_prev\_tmp.T) / batch\_size  
 db = np.sum(dZ, axis=1, keepdims=True) / batch\_size  
 dA\_prev = np.dot(W.T, dZ)  
 grads = [dW, db]  
  
 # Reshape dA\_prev to the shape of A\_prev  
 if len(A\_prev.shape) > 2:  
 dA\_prev = dA\_prev.T.reshape(self.input\_shape)  
  
 return dA\_prev, grads

در این قسمت تابع backward برای لایه fully connected پیاده‌سازی شده است. نیاز داریم که از مشتقات جزئی استفاده کنیم تا بتوانیم به دنبال وزن‌ها و بایاس‌های بهینه برای کاهش خطا باشیم.

برای این کار، ابتدا باید بررسی کنیم که آیا خروجی لایه قبل از نوع convolutional بوده است یا خیر. برای این منظور، از یک شرط if استفاده شده است. اگر خروجی قبلی کانولوشنی بود، باید ابعاد آن را به شکلی تغییر دهیم تا بتوانیم در محاسبات بعدی از آن استفاده کنیم.

سپس در بخش backward، باید با استفاده از مشتقات جزئی، مقدار gradient را برای وزن‌ها و بایاس‌ها محاسبه کنیم. پس از محاسبه این مقادیر، باید مشتق نسبت به فعالساز قبلی نیز محاسبه شود.

در نهایت مقدار‌های محاسبه شده برای gradient و مشتق نسبت به فعالساز قبلی در قالب لیستی به صورت [dW, db] و dA\_prev بازگردانده می‌شوند