



تمرین یک:

برای استفاده از الگوریتم های یادگیری در برنامه ها، استفاده از شیء گرای مناسب و مفید است (همانند آنچه در کتابخانه ی **scikit-learn** پیاده سازی شده). در این تمرین قصد داریم الگوریتم یادگیری پرسپترون را با رویکرد شیء گرای و ایجاد یک کلاس (**object oriented perceptron API**) پیاده سازی کنیم. در زمان استفاده از این کلاس، از آن شیء می گیریم و متد **fit** را برای آن با ارسال مجموعه داده های آموزشی فراخوانی می کنیم. این متد مقادیر مناسب برای وزن ها را (پارامترهای شیء ایجاد شده) می یابد و پس از آن فراخوانی متد پیش بینی (**predict**) به ازای یک ورودی، برچسب خروجی آن را پیش بینی می کند (دسته بندی دودویی). با توجه به فایل **perceptron.py** تمرین های زیر را انجام دهید.

الف. پیاده سازی متد **fit** را کامل کنید به نحوی که در آن به روزرسانی وزن ها به تعداد دفعات برابر با **self.n_iter** انجام شود.

ب. چرا تولید اعداد تصادفی به مقداردهی همه ی وزن ها با صفر ترجیح داده شده است؟ با توجه به پاسخ خود، نقش پارامتر **random_state** را شرح دهید.

ج. در فایل **Q1_3.py** به کمک کتابخانه ی **pandas**، مجموعه داده های **Iris** مربوط به ۱۵۰ نمونه آموزشی برای دسته بندی سه نوع گل با توجه به چهار ویژگی دانلود می شود. سپس تنها با در نظر گرفتن ویژگی های ردیف ۰ و ۲ (**sepal length** و **petal length**) برای دو کلاس **Iris-setosa** و **Iris-versicolor** مجموعه ی داده ی کوچکتری با قابلیت جداسازی خطی ایجاد می شود و با امکانات کتابخانه ی **matplotlib** به نمایش در می آید. برای این مجموعه ی داده ها، جداسازی را با کمک کلاس **perceptron** انجام دهید و نمودار تعداد خطا به ازای مرحله ی به روزرسانی رسم کنید. (فایل **Q1_3** را کامل کنید)

د. با توجه به حضور سه نوع کلاس داده در مجموعه داده های حاضر، راه حلی برای استفاده از **perceptron** در این مسئله ی چند کلاسه ی جدایی پذیر خطی بیان کنید.

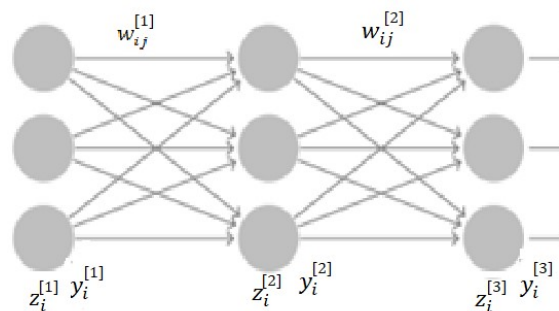
تمرین سری یک

تمرین دو:

کد مربوط به سوال قبل را به نحوی تغییر دهید (بازنویسی تابع fit و تغییرات جزئی دیگر) که به روزرسانی وزن ها با الگوریتم adaline و در هر تکرار با توجه به خطا روی کل نمونه های آموزشی انجام شود و کد را روی همان مجموعه داده ها آزمایش کنید.

تمرین سه:

الف. برای شبکه عصبی feed-forward زیر با وزن های نامگذاری شده در شکل، مشتقات خواسته شده را با قاعده ی زنجیره ای برای محاسبات error backpropagation بر حسب $\frac{\partial E}{\partial y_k^{[3]}}$ دست آورید (برای محاسبه ی هر عبارت می توانید از عبارات محاسبه شده ی قبلی استفاده کنید)



$$y_i^{[k]} = g(z_i^{[k]})$$

$$z_i^{[k]} = \sum_j w_{ij}^{[k-1]} y_j^{[k-1]}$$

$$\frac{\partial E}{\partial z_3^{[k]}} =$$

تمرین سری یک

$$\text{ب) } \frac{\partial E}{\partial z_2^{[k]}} =$$

$$\text{ج) } \frac{\partial E}{\partial w_{ij}^{[2]}} =$$

$$\text{د) } \frac{\partial E}{\partial w_{ij}^{[1]}} =$$

ب. توضیح دهید چرا تابع فعالیت لایه های مخفی باید غیر خطی باشد؟

تمرین چهار:

فضای فرضیه ها (hypothesis space) برای مسئله ی Regression در فضای ویژگی های دوبعدی را مجموعه ی مدل های به شکل زیر در نظر بگیرید:

$$y = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^2$$

فرض کنید نمونه های آموزشی به صورت $(x_1^{(i)}, x_2^{(i)}, y^{(i)})$ مشاهده شده اند. اگر هدف، کمینه سازی تابع خطای زیر باشد:

$$f(w_0, w_1, w_2, w_3) = \sum_{i=1}^n (y^{(i)} - w_0 - w_1(x_1^{(i)}) - w_2(x_2^{(i)}) - w_3(x_1^{(i)})^2)^2$$

گرایان تابع $f(w)$ نسبت به بردار پارامترها $(w = [w_0, w_1, w_2, w_3]^T)$ و با آن یک قاعده ی به روز رسانی ارائه کنید.

$$\nabla_w f(w) = \left[\frac{\partial f(w)}{\partial w_0}, \frac{\partial f(w)}{\partial w_1}, \frac{\partial f(w)}{\partial w_2}, \frac{\partial f(w)}{\partial w_3} \right]^T$$