

به نام خدا



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر



درس بینایی ماشین

تمرین شماره ۴

آذر ۱۳۹۹



فهرست

- سوال یک - فضای رنگی.....۳
- سوال دوم - شبکه عصبی چند لایه (MLNN).....۵
- سوال دوم - بخش امتیازی.....۷
- سوال سوم - شبکه های کانولوشنی (CNN).....۸



سوال یک - فضای رنگی

هدف این سوال آشنایی شما با انواع فضاهای رنگی و نحوه تبدیل این فضاها به یکدیگر است ، موارد خواسته شده را به ترتیب انجام داده و پاسخ آنها را گزارش نمائید.

الف) فضای رنگی XYZ را توضیح دهید ، چرا هیچ نمایشگری نمی تواند کل فضاهای XYZ را نمایش دهد؟

ب) تصویر شماره یک را به صورت RGB دریافت نموده و هر کانال آن را به صورت جداگانه Plot نمائید. تفاوت فضای رنگی RGB و sRGB در چیست؟

برای هر کدام از موارد زیر یک تابع بنویسید که با فراخوانی آن، تبدیل یا عمل مورد نظر انجام گیرد. دقت داشته باشید که نمی توانید از توابع آماده استفاده کنید.

ج) تصویر را با استفاده از یک تابع به فضای رنگی HSV برده و هر کانال آن را به صورت جداگانه Plot کنید، سپس توضیح دهید هر کانال نشان دهنده چه عاملی در این فضای رنگی است.

د) تصویر را با استفاده از یک تابع به فضای رنگی Y'CbCr برده و هر کانال آن را به صورت جداگانه Plot نمائید، سپس توضیح دهید هر کانال نشان دهنده چه عاملی در این فضای رنگی است.

و) تصویر را با استفاده از یک تابع به فضای Gray برده یا به عبارتی سیاه سفید کنید. توضیح دهید این کار را به چه صورتی انجام می دهید.

ه) یک تصویر جدید ایجاد کنید که فقط شامل چهره بازیگران تصویر ۱ باشد ، برای این کار ابتدا باید تصویر را از فضای RGB به فضای Normalized RGB برده (معادلات ۱ تا ۳) و سپس یک mask طراحی کنید (معادله ۴) و روی تصویر اعمال کرده و در نهایت آن را در تصویر ضرب کنید (معادله ۵) تا تنها چهره افراد حاضر در تصویر باقی بماند.

نکته : برای بردن به فضای Normalized RGB از معادلات زیر استفاده کنید.

$$r = \frac{R}{(R + G + B)} \quad (1)$$

$$g = \frac{G}{(R + G + B)} \quad (2)$$

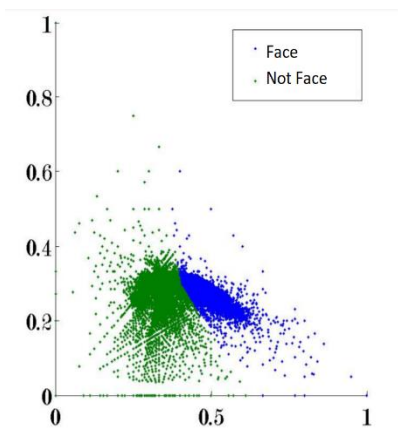
$$b = \frac{B}{(R + G + B)} \quad (3)$$

نکته: برای بدست آوردن Mask مورد نیاز، ابتدا کانال های r و g را شبیه تصویر ۲ پلات کرده و یک ناحیه از فضای r - g را پیدا کنید که شامل نقاط چهره باشد. یعنی یک تابع f پیدا کنید به گونه ای که پیکسل های چهره را با یک قاعده بتوان پیدا کرد. (این تابع را با آزمایش و خطا بر روی فضای r - g می توانید بدست آورید).

pixel (x,y) is a face pixel if $f(r(x,y), g(x,y)) \geq 0$

$$Mask(x,y) = \begin{cases} 1 & \text{if } f(r(x,y), g(x,y)) \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

$$J(x,y) = I(x,y)Mask(x,y) \quad (5)$$



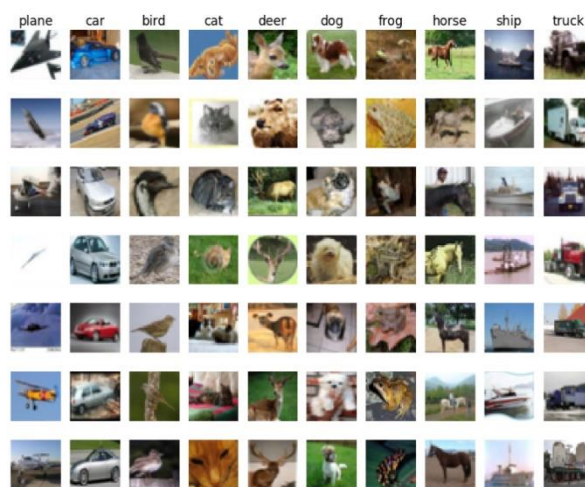
تصویر ۲) نمودار محورهای r و g



تصویر ۱)

سوال دوم- شبکه عصبی چند لایه (MLNN)

در این تمرین هدف پیاده سازی یک شبکه عصبی چندلایه (Multi-layer Neural Network) برای کاربرد طبقه بندی تصاویر (image classification) است. مجموعه داده cifar10 برای این تمرین انتخاب شده است که شامل 60000 تصویر رنگی با ابعاد 32×32 است که از ده کلاس متفاوت اشیا با سهم مساوی تشکیل شده است. این پایگاه شامل 5 دسته داده آموزش و یک دسته داده تست هر کدام شامل 10000 تصویر است. نمونه ای از این مجموعه داده را در تصویر ۳ مشاهده می کنید.



تصویر (۳) - نمونه مجموعه داده CIFAR-10

از طریق [این لینک](#) می توانید این مجموعه داده را دانلود نمایید.

با استفاده از کتابخانه Keras نیز می توانید مجموعه داده را دانلود نمایید.

```
from keras.datasets import cifar10

(x_train, y_train), (x_test, y_test) =
cifar10.load_data()
```

یک شبکه عصبی چند لایه برای طبقه بندی تصاویر این پایگاه داده پیاده سازی کنید که دارای یک مسیر forward برای محاسبه امتیاز کلاس های مختلف برای داده های ورودی باشد. سپس اختلاف امتیاز کلاس های تخمین زده شده با کلاس های حقیقی را به عنوان loss طبقه بندی در نظر بگیرید. در نهایت در مسیر backward با استفاده از قاعده error propagation برای آموزش وزن های شبکه عصبی خود طراحی کنید.



الف) در بخش اول شبکه را با مشخصات زیر پیاده سازی کنید:

✓ تعداد لایه های مخفی: ۱ لایه

✓ تعداد نورون های لایه مخفی: ۱۵۰ نورون

✓ تابع فعال ساز لایه مخفی: ReLU

در لایه آخر از تابع **Softmax** برای امتیاز دهی به کلاس های مختلف استفاده کنید. از تابع هزینه **Negative Log Likelihood** و بهینه ساز **SGD** همراه با **momentum** استفاده کنید و ضریب **momentum** را ۰.۹ در نظر بگیرید و نرخ یادگیری را در ابتدای یادگیری برابر با ۰.۰۰۱ در نظر بگیرید و در طول آموزش آنرا با گام های ثابت کاهش دهید. اندازه دسته (batch size) را برابر با ۱۲۸ در نظر بگیرید. از هر دسته از داده های آموزش مجموعه داده **CIFAR10**، حدود ۱۰٪ را برای **validation** و باقی را به عنوان داده **train** انتخاب کنید. داده های **test** نیز همان ۱۰ هزار تصویر مجموعه داده تست **CIFAR10** خواهند بود.

وزن های لایه مخفی را در ابتدا با ۰ مقدار دهی کنید. بردار تصاویر را پیش از ورودی به شبکه نرمال سازی (normalization) کنید.

نمودار های **loss** و **accuracy** را در طول آموزش بر حسب **epoch** برای داده های **train** و **validation** رسم کنید و آنها را تحلیل کنید. در تحلیل های خود مفهوم **overfit** را نیز بررسی کنید. معیار **accuracy** را برای داده های **test** گزارش کنید.

بخش های زیر را بر اساس مدل بخش "الف" که از آن با نام مدل پایه یاد میکنیم پیاده سازی کنید. در هر بخش با تغییر پارامتر مربوط به هر بخش و ثابت نگه داشتن باقی پارامتر ها، حساسیت مدل خود را نسبت به پارامتر متغیر خود بر روی داده های **validation** با رسم نمودار بررسی و تحلیل خود نسبت به تغییرات ایجاد شده را بیان کنید.

منحنی های **loss** و **accuracy** مدل را در موارد زیر برای مقایسه بررسی کنید. نرخ همگرایی را برای مدل ها به صورت تحلیلی مقایسه کنید و نیز معیار **accuracy** را بر روی داده گان تست در هر مدل گزارش کنید.

ب) شبکه را با تعداد ۲ و ۳ لایه مخفی پیاده سازی کنید و نتایج حاصل را با مدل پایه مقایسه و تحلیل کنید.



ج) تعداد نورون های لایه مخفی را به 50 و 200 نورون تغییر دهید و نتایج حاصل را با مدل پایه مقایسه و تحلیل کنید.

د) نرخ یادگیری اولیه را به 0.01 و 0.0001 تغییر دهید و نتایج حاصل را با مدل پایه مقایسه و تحلیل کنید.

و) مقدار دهی اولیه وزن ها را یکبار با توزیع نرمال با میانگین 0 و انحراف معیار 0.1 و یکبار با توزیع یکنواخت بین 0 و 1 مقدار دهی کنید و نتایج حاصل را با مدل پایه مقایسه و تحلیل کنید.

نکته : در تمام قسمت های سوال دوم و بخش امتیازی آن حق استفاده از توابع آماده یا کتابخانه های یادگیری ماشین را ندارید. (به جز داندلود مجموعه داده)

سوال دوم – بخش امتیازی

الف) تصاویر ورودی را پیش از ورودی به شبکه با آنالیز PCA به بردار های 128 تایی کاهش بعد دهید. نتایج حاصل را مانند قبل با مدل پایه مقایسه و تحلیل کنید.

ب) تصاویر ورودی را به کانال طیف خاکستری (gray scale) برده و سپس شبکه را با آنها آموزش دهید. چه تغییری مشاهده میکنید. نتایج خود را تحلیل کنید.

ج) بر اساس آنچه در درس بینایی ماشین آموخته اید و با تحلیل روی نمودار های بدست آمده از آموزش مدل پایه، چه پیشنهاد هایی برای بهبود عملکرد مدل نهایی خود دارید.



سوال سوم- شبکه های کانولوشنی (CNN)

این سوال در ادامه سوال سوم طرح شده است ، هدف در این تمرین افزایش دقت شبکه طبقه بند تصویر با اضافه کردن لایه های کانولوشنی و .. به شبکه است . در این سوال می توانید از کتابخانه های آماده استفاده کنید.

جهت یادآوری ، حتما موارد زیر را در گزارش برای قسمت های (الف) و (ب) بیاورید :

نمودار های **loss** و **accuracy** را در طول آموزش بر حسب **epoch** برای داده های **train** و **validation** رسم کنید و آنها را تحلیل کنید. معیار **accuracy** را نیز برای داده های **test** گزارش کنید.

الف) از ساختار شبکه بدست آمده در تمرین دوم استفاده نمائید و آن را مجددا پیاده سازی نمائید و این بار لایه های کانولوشنی به آن اضافه نمائید و شبکه را پیاده سازی نمائید. نتیجه بدست آمده را از نظر دقت و خطا با نتایج تمرین قبلی مقایسه نمائید.

ب) تعدادی از فیلترهای لایه های کانولوشنی را انتخاب کنید و روند تغییر آنها را در **epoch** های متفاوت **Visualize** کنید. چه تغییراتی مشاهده می کنید؟

ج) لایه **Pooling** را توضیح دهید و سپس این لایه ها را به توپولوژی شبکه اضافه نمائید و شبکه را پیاده سازی نمائید. نتیجه بدست آمده را از نظر دقت و خطا با معماری قسمت (الف) مقایسه نمائید.

نکته : می توانید از کتابخانه های یادگیری ماشین برای حل این سوال استفاده کنید.



نکات:

- مهلت تحویل این تمرین، پنج شنبه ۱۸ دی است.
- انجام این تمرین به صورت یک نفره می باشد.
- برای انجام تمرین ها فقط مجاز به استفاده از زبان های برنامه نویسی Python و MATLAB خواهید بود. در سؤالاتی که از شما خواسته شده است یک الگوریتم را پیاده سازی کنید **مجاز** به استفاده از توابع آماده **نمی** باشید **مگر** اینکه در صورت سوال اجازه استفاده از این توابع یا کتابخانه ها به شما داده شده باشد.
- داخل کدها کامنت های لازم را قرار دهید و تمامی موارد مورد نیاز برای اجرای صحیح کد را ارسال کنید.
- گزارش شما در فرآیند تصحیح از **اهمیت ویژه ای** برخوردار است. لطفاً تمامی نکات و فرض هایی که برای پیاده سازی ها و محاسبات خود در نظر می گیرید را در گزارش ذکر کنید.
- الزامی به ارائه توضیح جزئیات کد در گزارش نیست. اما باید نتایج بدست آمده را گزارش و تحلیل کنید.
- گزارش را در قالب تهیه شده که روی صفحه درس در سامانه Elearn بارگذاری شده، بنویسید. در صورت تمایل می توانید از Latex نیز برای نوشتن گزارش استفاده نمایید.
- در گزارش خود برای تصاویر زیرنویس و برای جداول هم بالانویس اضافه کنید.
- در صورت مشاهده **تقلب** نمرات تمامی افراد شرکت کننده در آن صفر لحاظ می شود.
- لطفاً گزارش ، فایل کدها و سایر ضmann مورد نیاز را با فرمت زیر در صفحه درس در سامانه یادگیری الکترونیکی بارگذاری نمایید.

HW4_[Lastname]_[StudentNumber].zip

- در صورت وجود هرگونه ابهام یا مشکل می توانید از طریق رایانامه زیر با دستیار آموزشی طراح تمرین (سوال ۱ و ۳ علی کریمی ، سوال ۲ پیمان باقرشاهی) در تماس باشید:

p.baghershahi@ut.ac.ir

alikarimi120@gmail.com