



دانشگاه صنعتی شریف - دانشکده برق

درس یادگیری ماشین

## تمرین کامپیوتری سری اول

مهلت تحویل: ۱۷ آذر ۱۴۰۲

نیمسال اول ۱۴۰۲-۱۴۰۳

---

آداب انجام تمرینات کامپیوتری

---

- **همفکری** دانشجویان در انجام تمرینات مانعی ندارد ولی تمرینات باید به صورت مشخص توسط خود تحویل دهنده انجام گرفته باشد. در زمان تصحیح تمرینات شباهت تمرینات بررسی خواهد شد و تمرین‌های با شباهت غیرعادی مورد قبول نیست.
- استفاده از منابع اینترنتی برای گرفتن ایده‌ی حل سوالات مانعی ندارد ولی کپی کردن پاسخ تمرینات از هر منبعی مورد قبول نیست.
- استفاده از کتابخانه‌های آماده در صورتی که در صورت تمرین قید نشده باشد، مانعی ندارد.
- کدهای تحویل شده باید قابلیت اجرای دوباره را داشته باشند. در صورت وجود مشکل در زمان تصحیح (حتی به دلیل خطای تایپی)، قسمت‌های غیرقابل اجرا مورد قبول نیست. لطفاً قبل از تحویل تمرین حتماً کد خود را بررسی فرمایید.
- فایل ZIP شامل کدهای تمرینات همراه گزارش (اگر در صورت تمرین نیاز به گزارش اعلام شده باشد) در CW آپلود گردد.

## مسئله اول

- ۱- ابتدا دیتاست breast cancer را از کتابخانه scikit-learn بخوانید و چند سطر اول آن را نمایش دهید.  
راهنمایی: از این کد استفاده کنید `from sklearn.datasets import load_breast_cancer`
- ۲- داده‌ها را به داده‌های train و test تقسیم کنید. از تقسیم ۳۰ و ۷۰ پیروی کنید.
- ۳- الگوریتم perceptron را پیاده سازی کرده و داده‌ها را طبقه‌بندی کنید. از الگوریتم sample mode باید استفاده کرد یا batch mode و چرا؟ صحت و confusion matrix را محاسبه کرده و نمایش دهید.
- ۴- تابع knn را پیاده سازی کنید. به ازای  $k = [1, 3, 5, 7]$  داده‌ها را طبقه‌بندی کنید. صحت و confusion matrix را برای همه آن‌ها محاسبه کرده و نمایش دهید.
- ۵- این بار داده‌ها را با svm طبقه‌بندی کرده و صحت و confusion matrix را محاسبه کنید.
- ۶- یک طبقه بند بیز پیاده سازی کنید و داده‌ها را طبقه‌بندی کنید. طبقه بند شما باید از توزیع گوسی برای توزیع احتمال هر کلاس استفاده کند. Confusion matrix و صحت را گزارش دهید.
- ۷- طبقه بندهای بالا را با هم مقایسه کنید کدام عملکرد بهتری دارد؟ آیا صحت معیار مناسبی برای این داده است؟ اگر نیست با معیار مناسب طبقه بندها را مقایسه کنید.
- شما مجاز از استفاده از تابع‌های آماده پایتون برای الگوریتم‌های perceptron و knn نیستید ولی استفاده از این توابع برای بدست آوردن confusion matrix و پیاده سازی svm ایرادی ندارد.

## مسئله دوم

- ۱- ابتدا 20,000 نمونه رندوم از توزیع گوسی با میانگین ۰ و واریانس ۱ بگیرید و در آرایه X بریزید.
- ۲- حال تابع پنجره پارزن را با کرنل uniform پیاده سازی کنید.
- ۳- به ازای  $\text{band width} = [0.01, 0.1, 1]$  تابع چگالی احتمال این توزیع را در بازه  $x = -4:4:1000$  تخمین بزنید.
- ۴- هیستوگرام X و توزیع گاوسی با میانگین صفر و واریانس یک و تخمین پنجره پارزن را در هر مرحله plot کنید.
- ۵- کدام band width مناسب‌تر است؟

- همراه با فایل ZIP کدها و خروجی، یک فایل گزارش شامل توضیحات در مورد آن چه انجام دادید و سوالات پرسش شده ارسال فرمایید. در صورت ارسال فایل نوتبوک می توانید توضیحات را در همان فایل بنویسید و نیازی به گزارش نیست.