



دانشکده‌ی مهندسی کامپیوتر

یادگیری ماشین

بهار ۱۴۰۰

تمرین سری ششم

مدرس: دکتر محمدحسین رهبان

زمان تحویل: ۳ خرداد

سوال ۱ Three Learning Principle

آ سه اصل یادگیری را نام برده و شرح کوتاهی از هریک ارائه دهید. (۲ نمره)

ب آیا پیچیده بودن مرز تصمیم‌گیری یک مدل در فضای داده، لزوماً به معنای نقض مفهوم *simplest* در اصل *Occam's Razor* است؟ با یک مثال

توضیح دهید. (۲ نمره)

پ رابطه بین *sampling bias* و *data snooping* را با ارائه یک مثال شرح دهید. (۲ نمره)

سوال ۲ Locally Weighted Linear Regression

آ مسئله *Locally Weighted Linear Regression* را با وزن‌های W در نظر بگیرید. ثابت کنید پارامترهای مدل از رابطه زیر بدست می‌آیند. (۳

نمره)

$$\theta = (X^T W X)^{-1} X^T W Y$$

ب وجود وزن‌های W چه کمکی به مسئله *Linear Regression* می‌کند؟ اگر وزن‌های W دارای توزیع گاوسی باشند، واریانس آن‌ها چه تاثیری بر

مدلی که یاد می‌گیریم، دارند؟ آیا لازم است واریانس این توزیع برای داده‌های مختلف مجموعه داده‌گان یکسان باشد؟ اثر متغییر بودن این واریانس‌ها

بر کارایی مدل را شرح دهید. (۳ نمره)

پ برای *adaptive* کردن W روش‌های گوناگونی پیشنهاد شده است. در یکی از این روش‌ها، واریانس مقادیر خروجی مدلی که یاد می‌گیریم به ازای

داده‌های ورودی در مجموعه داده‌گان کمینه می‌شود. تعریف‌های زیر را در نظر بگیرید:

$$\mu_x = \frac{\sum_i w_i x_i}{\sum_i w_i}, \quad \sigma_x^2 = \frac{\sum_i w_i (x_i - \mu_x)^2}{\sum_i w_i}, \quad \sigma_{xy} = \frac{\sum_i w_i (x_i - \mu_x)(y_i - \mu_y)}{\sum_i w_i}$$

$$\mu_y = \frac{\sum_i w_i y_i}{\sum_i w_i}, \quad \sigma_y^2 = \frac{\sum_i w_i (y_i - \mu_y)^2}{\sum_i w_i}, \quad \sigma_{y|x}^2 = \sigma_y^2 - \frac{\sigma_{xy}^2}{\sigma_x^2}$$

ثابت کنید واریانس این روش با رابطه زیر حساب می‌شود. افزایش یا کاهش پارامترهای توزیع گاوسی چه تاثیری بر این واریانس دارد؟ (۸ نمره)

$$\sigma_{\hat{y}}^2 = \frac{\sigma_{y|x}^2}{\left(\sum_i w_i\right)^2} \left(\sum_i w_i^2 + \frac{(x - \mu_x)^2}{\sigma_x^2} \sum_i w_i^2 \frac{(x_i - \mu_x)^2}{\sigma_x^2} \right)$$

سوال ۳ Gradient Boosting Algorithms

روش *Gradient Boost* از ترکیب خطی تعدادی مدل ضعیف برای ایجاد یک مدل قوی استفاده می‌کند. در واقع این روش تلاش می‌کند با یادگیری ضعیف چندین مدل و ترکیب آن‌ها به یک مدل آموزش دیده قوی برسد. هر مدل ضعیف برخی ضعف‌های مدل‌های قبلی خود، که هر داده ورودی به شکل سلسله‌مراتبی از آن‌ها عبور کرده تا به مدل ضعیف فعلی رسیده، را می‌پوشاند و مقدار تابع هزینه را به اندازه توان خودش کاهش می‌دهد. برای آشنایی بیشتر به منابع زیر مراجعه کنید:

- [Gradient Boosting](#)
- [Understanding Gradient Boosting Machines](#)
- [Chapter 12 Gradient Boosting](#)

آ مراحل مختلف الگوریتم را برای مسئله *Regression* و *Classification* شرح داده و تفاوت روند محاسبات آن‌ها را بیان کنید. (۲ نمره)

ب در مسئله *Classification* با استفاده از تابع هزینه *CrossEntropy*، مرحله محاسبه ضریب بهینه برای *Base Learner* جدید را بازنویسی کرده و مسئله بهینه‌سازی مربوطه را که به شکل زیر است، حل کنید. تعریف نمادهای این رابطه اینجا آمده است. (۸ نمره)

$$\gamma_m = \underset{\gamma \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$$

پ برای داده‌های جدول زیر، مسئله *Classification* را با استفاده از این الگوریتم تا ۳ مرحله جلو ببرید. *Base Learner* را *SVM* در نظر بگیرید. (۱۰ نمره)

<i>Likes Popcorn</i>	<i>Age</i>	<i>Favorite Color</i>	<i>Loves Troll 2</i>
<i>Yes</i>	<i>12</i>	<i>Blue</i>	<i>Yes</i>
<i>Yes</i>	<i>87</i>	<i>Green</i>	<i>Yes</i>
<i>No</i>	<i>44</i>	<i>Blue</i>	<i>No</i>
<i>Yes</i>	<i>19</i>	<i>Red</i>	<i>No</i>
<i>No</i>	<i>32</i>	<i>Green</i>	<i>Yes</i>
<i>No</i>	<i>14</i>	<i>Blue</i>	<i>Yes</i>

ت با استفاده از الگوریتم *XGBoost* به دسته‌بندی داده‌های جدول بخش پ پرداخته و روند محاسبات را مرحله به مرحله نشان دهید. (۱۰ نمره امتیازی)

سوال ۴ (عملی) با استفاده از روش‌های *XGBoost* و *Gradient Boost* به دسته‌بندی داده‌های *iris* در *sklearn.datasets* بپردازید. در این سوال ۱۰ درصد از داده‌ها را برای اعتبارسنجی^۱ و سایر داده‌ها را برای آموزش بکار بگیرید. برای هریک از این روش‌ها مقدار *Precision*، *Recall*، *F1 Score* و زمان اجرا روی داده‌های آموزش و اعتبارسنجی را محاسبه کرده و مقایسه کنید. همچنین *Confusion Matrix* را برای هر دو روش بدست آورید. (۲۵ نمره)

پاینده باشید

¹Validation