

بسمه تعالی

نیمسال دوم ۹۲-۹۱

مدرس: سلیمانی

نمره از ۱۰۰+۵

الگوشناسی آماری ۷۲۵-۴۰ (گروه ۲)

تمرین سری سوم: دسته‌بندی غیر خطی

موعد تحویل: ۸ اردیبهشت ۹۲

سوال ۱ (۱۴ نمره): تابع هسته (kernel)

۱.۱. اگر $k(x, x')$ یک هسته‌ی معتبر (valid) باشد:

a. (۲ نمره) نشان دهید به ازای هر تابع $f(\cdot)$ هسته‌ی $f(x)k(x, x')f(x')$ معتبر است.

b. (۲ نمره) همچنین نشان دهید هسته‌ی $\exp(k(x, x'))$ نیز هسته‌ای معتبر است.

c. (۳ نمره) با استفاده از بندهای a و b نشان دهید که هسته‌ی گاوسی هسته‌ی معتبری است.

۲.۱. (۳ نمره) فرض کنید $k(x, x') = \phi(x')^T \phi(x)$ و $k^2(x, x') = \psi(x')^T \psi(x)$ باشد. چه رابطه‌ای بین ابعاد ϕ و ψ برقرار است؟

۳.۱. (۴ نمره) یک فضای ویژگی اولیه d -بعدی در نظر بگیرید. اعمال یک هسته‌ی چندجمله‌ای درجه ۲ به صورت $k(x, x') = (x^T x' + 1)^2$ معادل با انتقال به یک فضای ویژگی جدید $\phi(x)$ خواهد شد که $k(x, x') = \phi(x')^T \phi(x)$ فضای ویژگی متناظر با هسته‌ی چندجمله‌ای درجه ۲ چندبعدی خواهد بود؟ برای هسته‌ی چندجمله‌ای درجه m $k(x, x') = (x^T x' + 1)^m$ این فضا چند بعدی است؟

سوال ۲ (۲۰ نمره): Kernel SVM

۱.۲. مجموعه داده‌ی موجود در شکل زیر را در نظر بگیرید:



- a. (۳ نمره) چنانچه برای دسته‌بندی از SVM حاشیه سخت (hard margin) با هسته‌ی چندجمله‌ای استفاده شود، حداقل چه درجه‌ای برای چندجمله‌ای پیشنهاد می‌نمایید؟ توضیح دهید.
- b. (۲ نمره) بردارهای پشتیبان (support vectors) را در بند a مشخص نمایید.
- c. (۳ نمره) رابطه‌ی مربوط به مرز حاصل را بر حسب x ، محل و ضرایب بردارهای پشتیبان مشخص شده در بند b تعیین کنید.

۲.۲. مجموعه داده‌ی موجود در بند قبل را در نظر بگیرید. فرض کنید پس از دریافت داده‌ها محدودده‌ی آنها در بازه‌ی $[0, 1]$ تنظیم شده است:

- a. (۳ نمره) اگر قصد داشته باشید از هسته‌ی گاوسی در SVM استفاده نمایید، از بین مجموعه مقادیر $\{0.001, 0.01, 0.1, 0.3, 0.5, 1\}$ برای پارامتر σ گاوسی کدام مقدار (یا مقادیر) را مناسب می‌دانید (توضیح).
- b. (۲ نمره) انتخاب پارامتر σ چه تاثیری در مرز حاصل و مقدار حاشیه خواهد داشت؟

- c. (۲ نمره) در حالت کلی آیا همیشه می‌توان با به‌کارگیری هسته‌ی گاوسی (با σ مناسب) در SVM بر روی هر مجموعه داده‌ای به خطای آموزش صفر رسید (توضیح)؟
- d. (۲ نمره) آیا در حالت کلی از بین مقادیر مختلف پارامتر σ ، مقداری که منجر به حاشیه‌ی بزرگتر می‌شود، مقدار بهتری است؟

۳.۲ (۳ نمره) فرض کنید برای جداسازی خطی یک مجموعه داده با استفاده از روش SVM از هسته $k(\mathbf{x}, \mathbf{x}')$ استفاده کرده‌ایم. حال اگر از هسته $ak(\mathbf{x}, \mathbf{x}') + b$ و ضرایب ثابت a و b مثبت است استفاده کنیم چه تغییری در مرز جدایی و میزان حاشیه خط جداساز و مقادیر ضرایب α_i ایجاد می‌شود؟

سوال ۳ (۱۶ نمره): مساله‌ی دوگان (dual problem)

۱.۳ (۵ نمره) مساله‌ی بهینه‌سازی SVM حاشیه نرم (soft margin) را در نظر بگیرید:

$$\min_{\mathbf{w}, w_0, \xi} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i$$

$$\text{s. t. } y^{(i)}(\mathbf{w}^T \mathbf{x}^{(i)} + w_0) \geq 1 - \xi_i \quad i = 1, \dots, N$$

نشان دهید مساله‌ی دوگان آن به صورت زیر است:

$$\max_{\alpha} \left\{ \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y^{(i)} y^{(j)} \mathbf{x}^{(i)T} \mathbf{x}^{(j)} \right\}$$

$$\text{s. t. } \sum_{i=1}^N \alpha_i y^{(i)} = 0$$

$$0 \leq \alpha_i \leq C, \quad i = 1, \dots, N$$

۲.۳ دوگان پرسپترون: قانون یادگیری پرسپترون که به صورت زیر است را در نظر بگیرید:

$$\mathbf{w}^{t+1} = \mathbf{w}^t + \eta \mathbf{x}^{(i)} y^{(i)}$$

- a. (۳ نمره) نشان دهید که در دسته‌بند پرسپترون، بردار وزن \mathbf{w} را می‌توان به صورت ترکیب خطی بردارهای $y^{(i)} \mathbf{x}^{(i)}$ نوشت ($y^{(i)} \in \{1, -1\}$). ضرایب α_i را در ترکیب خطی $\mathbf{w} = \sum_{i=1}^N \alpha_i y^{(i)} \mathbf{x}^{(i)}$ پیدا کنید.
- b. (۲ نمره) قانون یادگیری پرسپترون را برای به‌روزرسانی α_i (به جای \mathbf{w}) بازنویسی کنید.
- c. (۳ نمره) با استفاده از نتیجه‌ی به‌دست آمده در بند b، تعمیم مبتنی بر هسته پرسپترون (kernel-based Perceptron) که با استفاده از تابع هسته $k(\mathbf{x}, \mathbf{x}')$ بتواند قابلیت پیدا کردن مرزهای غیرخطی را داشته باشد، ارائه کنید.
- d. (۳ نمره) رابطه‌ی مربوط به مرز پرسپترون در حالت مبتنی بر هسته را بر حسب تابع هسته k و نمونه‌های آموزش و ضرایب آن‌ها مشخص نمایید.

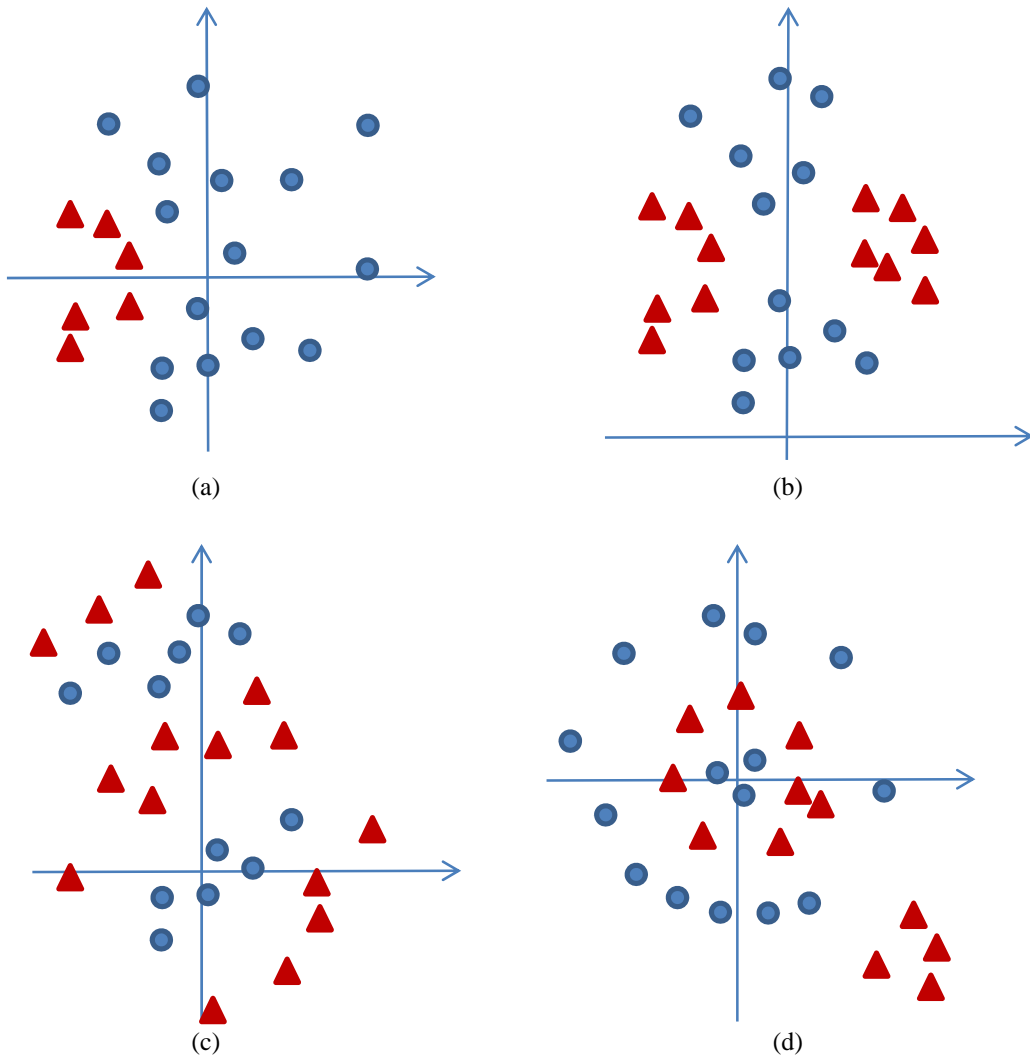
سوال ۴ (۲۵ نمره): مدل‌سازی توابع و تصمیم‌گیری با استفاده از شبکه عصبی

۱.۴ توابع بولین $\{T, F\}^d \rightarrow \{T, F\}$ را در نظر بگیرید. اگر از صفر برای نمایش F و از یک برای نمایش T استفاده کنیم:

a. (۵ نمره) آیا هر عبارت منطقی (متشکل از متغیرها و عملگرهای منطقی) بولین را می‌توان توسط یک شبکه عصبی با یک لایه مخفی که تابع آستانه نورون‌های لایه مخفی و خروجی آن پله واحد هستند، مدل کرد؟ توضیح دهید.

b. (۳ نمره) برای تابع NAND ساختار و وزن‌های یک شبکه با دو ورودی و یک خروجی که کارکرد NAND را دارد مشخص نمایید.

۲.۴. (۱۲ نمره) برای دسته‌بندی مجموعه داده‌های زیر شبکه‌های MLP با حداقل تعداد نورون و لایه‌ی ممکن ارائه کنید (تابع آستانه پله واحد در نظر گرفته شود) که قابلیت جداسازی کامل نمونه‌های دو دسته را داشته باشد.



۳.۴. (۵ نمره) یک شبکه عصبی طراحی نمایید که بتواند با دریافت یک بردار ویژگی d -بعدی مقدار تخمین غیر پارامتری چگالی احتمال را با استفاده از پنجره پارزن (Parzen window) در خروجی مشخص کند. فرض کنید نمونه‌های موجود برای محاسبه‌ی تخمین $\{x^{(1)}, \dots, x^{(N)}\}$ باشد. برای طراحی باید معماری شبکه، ورودی‌ها، خروجی، اتصالات و تابع فعالیت نورون‌ها را مشخص نمایید.

سوال ۵ (۳۰ نمره): پیاده‌سازی دسته‌بندی‌های غیرخطی

مجموعه داده‌ی MNIST شامل ارقام دست‌نویس که در فایل "MNIST.mat" در اختیار شما قرار گرفته، را در نظر بگیرید. هر داده در این مجموعه یک تصویر 28×28 است. هدف آن است که به ازای تصویر 28×28 ورودی دسته‌بند تشخیص دهد که تصویر مربوط به کدام رقم است. هر سطر داده موجود در فایل داده MNIST حاصل برداری کردن یک تصویر 28×28 است که در انتهای آن مشخص شده است که این تصویر مربوط به کدام یک از ارقام 0 تا 9 بوده است (بنابراین از ستون‌های 1 تا 784 برای ویژگی‌های داده و از ستون 785 برای برچسب استفاده نمایید). از 4000 داده‌ی اول برای آموزش و از مابقی داده‌ها برای آزمون دسته‌بند استفاده شود.

۱.۵. شبکه عصبی MLP با الگوریتم آموزش پس‌انتشار خطا (Error Backpropagation):

- a. (۵ نمره) برای پیاده‌سازی شبکه عصبی از کد اولیه موجود در فایل "NN.zip" که در اختیار شما قرار گرفته، استفاده کنید. برای این‌که از یک شبکه با $K = 10$ خروجی برای دسته‌بندی ورودی‌ها استفاده نمایید، ابتدا باید برچسب‌های داده را به صورت بردار خروجی مربوطه کد نمایید (مثال: برچسب مشخص‌کننده‌ی رقم 3 به بردار $[0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0]$ تبدیل شود). یک شبکه‌ی MLP با یک لایه‌ی مخفی شامل 100 نورون در نظر بگیرید، پس از آموزش شبکه، درصد خطای آن را روی داده‌ی آموزش و آزمون مشخص نمایید.
- b. (۷ نمره) با استفاده از 5-fold CV تعداد نورون‌های لایه‌ی مخفی MLP را از بین مجموعه $\{5, 10, 20, 50, 100, 200, 500, 1000\}$ انتخاب نمایید و درصد خطای شبکه‌ی حاصل را روی نمونه‌های آموزش، اعتبارسنجی و آزمون گزارش نمایید. همچنین نمودار MSE را روی نمونه‌های آموزش و آزمون به ازای تغییر تعداد نورون لایه‌ی مخفی رسم کنید.

۲.۵. Kernel SVM (one vs. all): برای پیاده‌سازی SVM موردنظر می‌توانید از بسته‌ی بهینه‌سازی محدب 'cvx' یا از 'libsvm' استفاده نمایید.

- a. (۶ نمره) SVM حاشیه نرم (soft-margin) با هسته‌ی خطی $k(x, x') = x^T x'$ را که قابلیت دسته‌بندی چند دسته‌ای را داشته باشد، پیاده‌سازی کنید. مقادیر پارامتر C را از بین مجموعه $\{0.001, 0.01, 0.1, 0.5, 1, 5, 10\}$ با استفاده از 5-fold CV تعیین نمایید. درصد خطای دسته‌بند حاصل را روی نمونه‌های آموزش و آزمون گزارش نمایید.
- b. (۷ نمره) SVM حاشیه نرم (soft-margin) با هسته‌ی گاوسی و قابلیت دسته‌بندی چند دسته‌ای را مشابهاً پیاده‌سازی نمایید. مقادیر پارامتر C و پهنای پنجره σ را به ترتیب از بین مجموعه‌های $\{0.001, 0.01, 0.1, 0.5, 1, 5, 10\}$ و $\{0.001, 0.01, 0.1, 1, 10, 100\}$ با استفاده از 5-fold CV تعیین نمایید. درصد خطای دسته‌بند حاصل را روی نمونه‌های آموزش، اعتبارسنجی و آزمون گزارش نمایید.

۳.۵. (۵ نمره) نتایج به‌دست آمده در بندهای مختلف برای MLP و Kernel SVM را مقایسه و تحلیل کنید.

¹ <http://cvxr.com/cvx/>

² www.csie.ntu.edu.tw/~cjlin/libsvm