



- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص شده است.
- همکاری و همفکری شما در انجام تمرین مانعی ندارد اما پاسخ ارسالی هر کس حتما باید توسط خود او نوشته شده باشد.
- در صورت همفکری و یا استفاده از هر منابع خارج درسی، نام همفکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید.
- لطفا تصویری واضح از پاسخ سوالات نظری بارگذاری کنید. در غیر این صورت پاسخ شما تصحیح نخواهد شد.
- موضوعات تمرین: شبکه-SVM های عصبی

سوالات نظری (۵۰ نمره)

۱. (۲۰ نمره) در کلاس درس درباره استفاده از SVM ها برای دسته‌بندی بحث شد. حال در این سوال قصد داریم این مفهوم را به مسئله رگرسیون انتقال دهیم. برای این منظور، همان روالی که انجام شد را مرحله به مرحله انجام می‌دهیم. فرض کنید داده‌هایتان $(x^{(1)}, y^{(1)}) \dots (x^{(n)}, y^{(n)})$ باشند که $x^{(i)} \in \mathbb{R}^d, y^{(i)} \in \mathbb{R}$ ، یک تابع ضرر متداول برای این منظور به صورت

$$L_\epsilon(x, y, f) = |y - f(x)|_\epsilon = \max(\cdot, |y - f(x)| - \epsilon)$$

است که با بکارگیری آن، تابع هزینه کلی به صورت زیر درمی‌آید.

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^n L_\epsilon(x^{(i)}, y^{(i)}, f)$$

الف) با تعریف متغیر slack (ξ_i^*, ξ_i) و اعمال شرط مناسب روی آن، نشان دهید صورت اصلی (primal) این مسئله (که یک مسئله quadratic است) به شکل

$$\min_{w \in \mathbb{R}^m, \xi \in \mathbb{R}^n, \xi^* \in \mathbb{R}^n} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i^* + \xi_i)$$

می‌باشد. (راهنمایی: همانطور که خواندید، در soft margin SVM متغیر ξ_i میزان تخطی حاشیه را نشان می‌داد. در این مسئله ξ_i را میزان تخطی مربوط به بیشتر بودن پیشبینی از y_i و ξ_i^* را میزان تخطی کمتر بودن پیشبینی از y_i بگیرید.)

ب) ابتدا تابع لاگراژین را برای صورت اصلی بنویسید سپس با عوض کردن ترتیب max و min و استفاده از شرط‌های K.K.T به صورت دوگان برسید.

$$\max_{\alpha \in \mathbb{R}^n, \alpha^* \in \mathbb{R}^n} -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle - \epsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*)$$

با شرط

$$\alpha_i, \alpha_i^* \in [\cdot, C].$$

- پ) توضیح دهید که آیا صورت دوگان مسئله با solver optimization quadratic قابل حل است؟
- ت) در این مسئله support vector ها به چه صورت مشخص می‌شوند؟
- ث) رابطه‌ای برای بیشینه داده‌ی جدید بنویسید و توضیح دهید که آیا می‌شود از تکنیک کرنل استفاده کرد؟
- ج) تغییر ϵ موجب چه می‌شود؟ تغییر C چگونه؟

۲. (۱۰ نمره) حداقل تعداد نمونه‌های m مورد نیاز را تعیین کنید تا تضمین شود که، با حداقل اطمینان ۹۵٪، فرضیه آموخته شده h دارای نرخ خطای کمتر از ۵٪ در داده‌های نادیده خواهد بود، با فرض حالت قابل تحقق (به عنوان مثال، یک فرضیه در H وجود دارد که مفهوم هدف را کاملاً توصیف می‌کند).
داده‌های داده شده:

$$|H| = 1000$$

$$\epsilon = 0.05$$

$$\delta = 0.05$$

۳. (۲۰ نمره الف) یک شبکه عصبی با ورودی x در نظر بگیرید. برای بدست آوردن خروجی محاسبات زیر بر روی x انجام می‌شود.

$$z = wx + b$$

$$y = \sigma(z)$$

$$\mathcal{L} = \frac{1}{2}(y - t)^2$$

$$\mathcal{R} = \frac{1}{2}w^2$$

$$\mathcal{L}_{\text{reg}} = \mathcal{L} + \lambda \mathcal{R}$$

گراف محاسباتی این مسئله را رسم کنید و مشتقات \mathcal{L}_{reg} را آپدیت کنید.

ب) پارامترهای یک شبکه عصبی در ابتدا به صورت تصادفی و با مقادیر کوچک مقداردهی می‌شوند. توضیح دهید در صورت عدم رعایت این دو ویژگی در مقداردهی چه مشکلاتی بروز پیدا می‌کند.

ج) وزن‌های شبکه عصبی بدست آمده از قسمت اول سوال را با مقادیر تصادفی دلخواه مقداردهی کنید و برای یک ورودی دلخواه، با توجه به مشتقاتی که در قسمت اول بدست آوردید، با اعمال بهینه‌سازی descent gradient برای یک اپاک که با نرخ یادگیری ۰.۱ وزن‌های شبکه را آپدیت کنید.

سوالات عملی (۵۰ نمره)

به نوتبوک های ضمیمه شده تمرین مراجعه کنید.