Lagrange Multipliers Recap

Problem $f: \mathbb{R} \to \mathbb{R}$

minimize $f(x)$

subject to equality constraint $g(x) = 0$

The associated Lagrangian is

$$\mathcal{L}(x, \lambda) = f(x) - \lambda g(x)$$

Problem is optimized when

$$0 = \frac{\partial \mathcal{L}}{\partial x} = f'(x) - \lambda g'(x) \Rightarrow \frac{\partial}{\partial x} f = \lambda \frac{\partial}{\partial x} g$$

$$0 = \frac{\partial \mathcal{L}}{\partial \lambda} = -g(x) \Rightarrow g(x) = 0$$

KKT extend Lagrange multipliers to allow for inequality constraints! $g(x) \leq 0$

★ These give solutions to problems where we might not be able to parametrize the exact values!

Recall from last week

Dual SVM Algorithm

Input: $(x_1, y_1), \dots, (x_n, y_n) \in \mathbb{R}^D \times \{\pm 1\}$

- Kernel map $K(x, y)$ } Choosing good Kernels is an active area of research!
- Regularization parameter C.

Main Step Solve Dual SVM optimization

$$\max_{\alpha \in \mathbb{R}^n} \left\{ -u^T G u + \sum_{k=1}^{n} \alpha_k \;\middle|\; \begin{array}{c} \sum \alpha_k y_k = 0 \\ 0 \leq \alpha_i \leq C \end{array} \right\}$$

$$u = \left(\tfrac{1}{2} \alpha_k y_k \right)_{k=1}^{n} \qquad G = \left(K(x_i, x_j) \right)_{i,j=1}^{n}$$

Output Solution of Dual SVM. $f_\theta(x) = \mathrm{Sgn}(\langle \theta, x \rangle + \theta_0)$

Today

- Derive Dual SVM by adding parameters to Soft Margin SVM.
- See how output is found.

Lagrange function for Soft margin SVM.

$$\mathcal{L}(\theta', \theta_0, \xi, \alpha, \beta) = \|\theta'\|^2 + \sum_{k=1}^{n} \xi_k (C - \beta_k)$$
$$- \sum_{k=1}^{n} \alpha_k \left(y_k(\langle\theta', x_k\rangle + \theta_0) - (1 - \xi_k)\right)$$

Soft margin SVM

$\alpha_k, b_k > 0$, in Soft margin SVM, $y_k(\langle\theta', x_k\rangle + \theta_0) \geq 1 - \xi_k$

$\boxed{\text{Karush-Kuhn-Tucker Conditions}}$ (KKT)

↳ Give conditions for when we can solve an optimization problem. We want to find

$$\min_{\theta, \xi} \max_{\alpha, \beta} \mathcal{L}(\theta', \theta_0, \xi, \alpha, \beta).$$

By KKT, optimum occurs when

$$\text{✪} \quad \underset{①}{\frac{\partial \mathcal{L}}{\partial \theta'} = 0}, \quad \underset{②}{\frac{\partial \mathcal{L}}{\partial \theta_0} = 0}, \quad \underset{③}{\frac{\partial \mathcal{L}}{\partial \xi} = 0}$$

Exercise

- Compute + simplify ⊛, don't use lecture notes since they make it fancy looking.
- Plug solution to ⊛ into $\mathcal{L}$ to simplify

---

① $\underbrace{\frac{\partial \mathcal{L}}{\partial \theta'}}_{} = \left[\frac{\partial \mathcal{L}}{\partial \theta_j} \right]_{j=1}^{n} = \left[\frac{\partial}{\partial \theta_j} \left(\sum_{\ell=1}^{D} \theta_\ell^2 - \sum_{k=1}^{n} \alpha_k y_k \left(\sum_{\ell=1}^{n} \theta_\ell x_k^{(\ell)} \right) \right) \right]$

$$= \left[2\theta_j - \sum_{k=1}^{n} \alpha_k y_k \sum_{\ell=1}^{D} \frac{\partial}{\partial \theta_j} (\theta_\ell) x_k^{(\ell)} \right]_{j=1}^{n}$$

$$= \left[2\theta_j - \sum_{k=1}^{n} \alpha_k y_k \, x_k^{(j)} \right]_{j=1}^{n}$$

$$= 2\theta' - \sum_{k=1}^{n} \alpha_k y_k \, x_k = 0$$

The x_k with $\alpha_k = 0$ are the support of the vector θ'! hence SVM.

$$\theta' = \frac{1}{2} \sum_{k=1}^{n} \alpha_k y_k x_k$$

* From last week $\psi(x) = \frac{1}{2} \sum_{k=1}^{n} \alpha_k y_k K(x_k, x)$

$$= \langle \theta', x \rangle$$

② $\frac{\partial \mathcal{L}}{\partial \theta_0} = \frac{\partial}{\partial \theta_0}\left[-\sum_{k=1}^{n} \alpha_k y_k \theta_0\right] = -\sum_{k=1}^{n} \alpha_k y_k$

$$\Rightarrow \boxed{\sum_{k=1}^{n} \alpha_k y_k = 0}$$

③ $\frac{\partial \mathcal{L}}{\partial \xi} = \left[\frac{\partial \mathcal{L}}{\partial \xi_j}\right]_{j=1}^{n} = \left[\frac{\partial}{\partial \xi_j}\left[\sum_{k=1}^{n}\left[\xi_k (C-\beta_k) + \alpha_k(1-\xi_k)\right]\right]\right]_{j=1}$

$$= \left[C - \beta_j - \alpha_j\right]_{j=1}^{n} = \left[C - (\beta_j + \alpha_j)\right]_{j=1}^{n} = 0$$

$$\Rightarrow \forall\ 1 \le j \le n \quad \beta_j + \alpha_j = C$$

$$\Rightarrow \alpha_j \le C$$

$\boxed{\text{Last step of exercise ③}}$ + ②

plug in $\beta_j + \alpha_j = C$ to set $C - \beta_j = \alpha_k$

$$\theta' = \frac{1}{2}\sum_{k=1}^{n} \alpha_k y_k x_k \qquad \sum \alpha_k y_k = 0$$

$$\mathcal{L}(\theta', \theta_0, \xi, \alpha, \beta) = \underline{\|\theta'\|^2} + \sum_{k=1}^{n} \xi_k (C - \beta_k)$$

$$- \sum_{k=1}^{n} \alpha_k \left(\underline{y_k(\langle \theta', x_k \rangle + \theta_0)} - (1 - \xi_k) \right)$$

$$= \|\theta'\|^2 + \sum_{k=1}^{n} \xi_k \alpha_k$$

$$- \sum_{k=1}^{n} \alpha_k (y_k \langle \theta', x_k \rangle) + \theta_0 \alpha_k y_k - \alpha_k + \alpha_k \xi_k$$

So optimal occurs with finding

$$\max_{\alpha} \left\{ \|\theta'\|^2 + \sum_{k=1}^{n} \alpha_k - \sum_{k=1}^{n} \alpha_k (y_k \langle \theta', x_k \rangle) \right\}$$

Subject to $0 \leq \alpha_k \leq C, \sum \alpha_k y_k = 0$

Plugging in $\theta' \stackrel{①}{=} \frac{1}{2} \sum \alpha_k y_k x_k$,

$$\langle \theta', x_k \rangle = \frac{1}{2} \sum_{j=1}^{n} \alpha_j y_j \langle x_j, x_k \rangle$$

$$\|\theta'\|^2 = \langle \theta', \theta' \rangle = \frac{1}{4} \sum_{k=1}^{n} \sum_{j=1}^{n} \alpha_k \alpha_j y_k y_j \langle x_k, x_j \rangle$$

$$\star\star = -\frac{1}{4}\sum_{j=1}^{n}\sum_{k=1}^{n}\alpha_k\alpha_j y_k y_j \langle x_k, x_j\rangle$$

$$u = \left(\frac{1}{2}\alpha_k y_k\right)_{k=1}^{n}$$

$$G = \left(\langle x_i, x_j\rangle\right)_{i,j=1}^{n}$$

Notice

$$u^T G u = \left[\frac{1}{2}\alpha_1 y_1 \cdots \frac{1}{2}\alpha_k y_k\right]\begin{bmatrix}\langle x_1, x_1\rangle & \cdots & \langle x_n, x_1\rangle \\ \vdots & & \vdots \\ \langle x_n, x_1\rangle & & \langle x_n, x_n\rangle\end{bmatrix} u$$

$$= \left[\frac{1}{2}\sum_{k=1}^{n}\alpha_k y_k\langle x_k, x_1\rangle \cdots \frac{1}{2}\sum_{k=1}^{n}\alpha_k y_k\langle x_k, x_n\rangle\right]\begin{bmatrix}\frac{1}{2}\alpha_1 y_1 \\ \vdots \\ \frac{1}{2}\alpha_n y_n\end{bmatrix}$$

$$= \frac{1}{4}\sum_{j=1}^{n}\sum_{k=1}^{n}\alpha_k\alpha_j y_j y_k\langle x_k, x_j\rangle \checkmark$$