ML1 Home Assignment 5

id: 12179078

Volodymyr Medentsiy

№1

$D = \{x_1, \dots, x_N\}$ with $x_i \in \{1, \dots, K\}$ – roll of a dice.

$$p(D|\theta) = \prod_{k=1}^{K} \theta_k^{N_k}, \text{ with } N_k = \sum_{n=1}^{N} [x_n = k]$$

$$\theta_k: \quad \sum_{k=1}^{K} \theta_k = 1$$
$$\theta_k \geq 0 \quad \forall k.$$

Prior on θ: $\quad p(\theta|\alpha) = Dir(\theta|\alpha) = \frac{1}{B(\alpha)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1}$

Posterior: $p(\theta|D) \propto p(D|\theta) \cdot p(\theta) \propto \prod_{k=1}^{K} \theta_k^{N_k + \alpha_k - 1}$

So $p(\theta|D) = Dir(\theta | \alpha_1 + N_1, \dots, \alpha_K + N_K)$

1.1. Derive the log-posterior:

$$\log p(\theta|D) = \sum_{k=1}^{K} (N_k + \alpha_k - 1) \log \theta_k$$

1.2. Define the Lagrangian:

$$\mathcal{L}(\theta, \lambda, \mu) = \sum_{k=1}^{K} (N_k + \alpha_k - 1) \log \theta_k + \lambda\left(\sum_{k=1}^{K} \theta_k - 1\right) +$$
$$+ \sum_{k=1}^{K} \mu_k \theta_k.$$

1.3. KKT conditions. – $3K+1$ conditions:

$$\sum_{k=1}^{K} \theta_k - 1 = 0 \qquad (1)$$
$$\mu_k \geq 0 \qquad (2)$$
$$\mu_k \theta_k = 0 \qquad (3)$$
$$\theta_k \geq 0 \qquad (4)$$

1.4. Find θ_{MAP}:

$$\frac{dl}{d\theta_k} = \frac{N_k + \alpha_k - 1}{\theta_k} + \lambda + \mu_k = 0 \Rightarrow$$

$$\Rightarrow \theta_k = \frac{1 - N_k - \alpha_k}{\mu_k + \lambda}$$

and: $\underbrace{\sum_{k=1}^{K}(N_k + \alpha_k - 1)}_{N-K+\sum_{k=1}^{K}\alpha_k} + \underbrace{\sum_{k=1}^{K}\theta_k \cdot \lambda}_{\lambda \text{ using}(1)} + \underbrace{\sum_{k=1}^{K}\theta_k \mu_k}_{\overset{||}{0} \text{ using } (3)} = 0$

So $\lambda = K - N - \sum_{k=1}^{K}\alpha_k$ and $\theta_k = \frac{1 - N_k - \alpha_k}{\mu_k + K - N - \sum_{k=1}^{K}\alpha_k}$

Using (3) $\left[\begin{array}{l}\theta_k = 0 \\ \mu_k = 0\end{array}\right.$. $\theta_k = 0 \Leftrightarrow N_k + \alpha_k = 1$ and μ_k could be of any choice in this case; if $\mu_k = 0$ then

$\theta_k = \frac{1 - N_k - \alpha_k}{K - N - \sum_{k=1}^{K}\alpha_k}$.

In general case we can state that $\theta_k^{MAP} = \frac{1 - N_k - \alpha_k}{K - N - \sum_{k=1}^{K}\alpha_k}$.

W2

2.1. Prime Lagrangian:

$$\ell(R, \alpha, \xi, \lambda, \mu) = \frac{1}{2}\alpha^2 + C\sum_{n=1}^{N}\xi_n - \sum_{n=1}^{N}\lambda_n\left(t_n(\alpha\|x_n\| - R) - 1 + \xi_n\right)$$
$$- \sum_{n=1}^{N}\mu_n\xi_n$$

2.2. KKT conditions

$\forall n$:

$$\lambda_n \geq 0 \qquad \mu_n \geq 0$$
$$\lambda_n \cdot (t_n(\alpha\|x_n\| - R) - 1 + \xi_n) = 0 \qquad \mu_n\xi_n = 0$$
$$(t_n(\alpha\|x_n\| - R) - 1 + \xi_n) \geq 0 \qquad \xi_n \geq 0$$

$6N$ KKT conditions

2.3. Dual Lagrangian:

$$\frac{d\ell}{dR} = \sum_{n=1}^{N}\lambda_n t_n = 0$$

$$\frac{d\ell}{d\alpha} = \alpha - \sum_{n=1}^{N}\lambda_n t_n\|x_n\| = 0 \Rightarrow \alpha = \sum_{n=1}^{N}\lambda_n t_n\|x_n\|$$

$$\frac{d\ell}{d\xi_n} = C - \lambda_n - \mu_n = 0 \Rightarrow \lambda_n + \mu_n = C$$

$$\tilde{\ell}(\lambda, \mu) = \frac{1}{2}\left(\sum_{n=1}^{N}\lambda_n t_n\|x_n\|\right)^2 + C\sum_{n=1}^{N}\xi_n - \sum_{n=1}^{N}\lambda_n t_n(\alpha\|x_n\| - R) +$$
$$+ \sum_{n=1}^{N}\lambda_n - \sum_{n=1}^{N}\underbrace{(\lambda_n + \mu_n)}_{C}\xi_n = -\frac{1}{2}\left(\sum_{n=1}^{N}\lambda_n t_n\|x_n\|\right)^2 + \sum_{n=1}^{N}\lambda_n$$

with $\lambda_n \in [0; C]$
$$\sum_{n=1}^{N}\lambda_n t_n = 0$$

2.4. The explicit form of kernel:

$$-\frac{1}{2}\left(\sum_{n=1}^{N} \lambda_n t_n \|x_n\|\right)^2 + \sum_{n=1}^{N} \lambda_n = -\frac{1}{2} \sum_{n=1}^{N}\sum_{m=1}^{N} \lambda_n t_n \|x_n\| \|x_m\| t_m \lambda_m +$$

$$+ \sum_{n=1}^{N} \lambda_n = -\frac{1}{2} \sum_{n=1}^{N}\sum_{m=1}^{N} \lambda_n t_n \kappa(x_n, x_m) t_m \lambda_m + \sum_{n=1}^{N} \lambda_n,$$

So $\kappa(x_n, x_m) = \|x_n\| \cdot \|x_m\|$.

2.5.

$$\lambda_n \in (0; C) \Rightarrow \mu_n > 0 \Rightarrow \begin{cases} \xi_n = 0 \\ t_n(\alpha \|x_n\| - R) - 1 + \xi_n = 0 \end{cases} \Rightarrow$$

$\Rightarrow x_n$ on the decision boundary. (x_n – is the suport vector) $\Rightarrow$

$\Rightarrow$ minimum 2 λ_n: $\lambda_n \in (0; C)$, because there should be at least 2 suport vectors to define a decision boundary (one support vector for each circle).

2.6. Let R^*, α^* be the solutions of the dual program.

Then for new x^*:

$$t^* = 1 \Leftrightarrow \alpha^* \|x^*\| \le R^* \Leftrightarrow \sum_{n=1}^{N} \lambda_n^* t_n \underbrace{\|x_n\| \|x^*\|}_{\kappa(x_n, x^*)} = \sum_{n=1}^{N} \lambda_n^* t_n \kappa(x_n, x^*) \le R^*$$

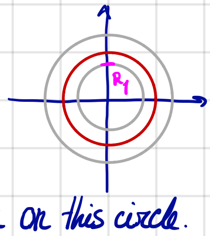
$$t^* = -1 \Leftrightarrow \alpha^* \|x^*\| > R^* \Leftrightarrow \sum_{n=1}^{N} \lambda_n^* t_n \|x_n\| \|x^*\| = \sum_{n=1}^{N} \lambda_n^* t_n \kappa(x_n, x^*) > R^*$$

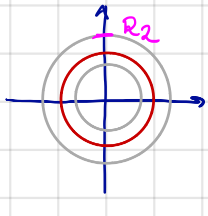
2.7.

$$\lambda_n > 0 \Rightarrow t_n(\alpha \|x_n\| - R) - 1 + \xi_n = 0 \Rightarrow$$

$\Rightarrow x_n$ is on the decision boundary or in the wrong area of it.

if $t_n = -1$, x_n will be out of the circle with radius R_1 or on this circle.

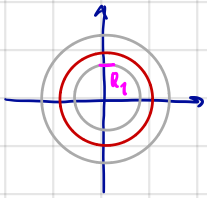




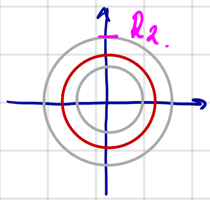
if $t_n = 1$, then x_n is in the circle of radius R_2 or on it.

$\mu_n > 0 \Rightarrow \xi_n = 0 \Rightarrow x_n$ is on or in the right area of the circle. So

If $t_n = -1$, x_n is on or in the circle of radius R_1.



If $t_n = 1$, then x_n is on or out of the circle of radius R_2.



2.8. $\mu_i^* = C - \lambda_i^*$

$$\alpha^* = \sum_{n=1}^{N} \lambda_n^* t_n \|x_n\|.$$

$\lambda_i^* \cdot (t_i(\alpha^* \|x_i\| - R^*) - 1 + \xi_i^*) = 0$, when $\mu_i^* \in (0; C)$: $\xi_i^* = 0$.

and then $\lambda_i^* (t_i(\alpha^* \|x_i\| - R^*) - 1) = 0 \Rightarrow R^* = \alpha^* \|x_i\| - \frac{1}{t_i} =$

$$= \sum_{n=1}^{N} \lambda_n^* t_n \|x_n\| \|x_i\| - \frac{1}{t_i} = \sum_{n=1}^{N} \lambda_n^* t_n K(x_n, x_i) - \frac{1}{t_i}$$

So: $\xi_j^* = 1 - t_j(\alpha^* \|x_j\| - R^*)$ for $\mu_j^* = 0$

$\xi_j^* = 0$ for $\mu_j^* > 0$

2.9. When using RBF kernel the decision boundary could differ from circle. It will have circle-line shape but could have a different radius in each point, so we can encounter such decision boundary