参考答案

① 要分类的数据集线性不可分，因此必须使用 Kernel

$$K(x^{(i)}, x^{(j)}) = \phi(x^{(i)})^T \phi(x^{(j)})$$

因为 Kernel Trick，我们不需要显式定义 $\phi(x^{(i)})$，只要写出 $K(x^{(i)}, x^{(j)})$ 即可。可以使用常用的 kernel 比如 RBF Kernel

$$K(x, z) = \exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right)$$

，也可以用其他 kernel，但若使用不常见的 kernel 或通过 $\phi(x^{(i)})$ 来得到 kernel 的，必须证明 kernel 的合理性（满足两个条件：对称且半正定）

② 建立目标函数：

也可写成 $\phi(x^{(i)})$，同下

$$\min_{w,b} \frac{1}{2}\|w\|^2$$

$$\text{s.t.} \ -y^{(i)}(w^T \boxed{x^{(i)}} + b) + 1 \le 0, \ i=1, \dots, m$$

Lagrangian：

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \alpha_i [y^{(i)}(w^T \boxed{x^{(i)}} + b) - 1]$$

$$\frac{\partial L(w,b,\alpha)}{\partial w} = w - \sum_{i=1}^{m} \alpha_i y^{(i)} \boxed{x^{(i)}} = 0 \Rightarrow w = \sum_{i=1}^{m} \alpha_i y^{(i)} \boxed{x^{(i)}}$$

$$\frac{\partial L(w,b,\alpha)}{\partial b} = \sum_{i=1}^{m} \alpha_i y^{(i)} = 0$$

$$\Rightarrow L(w, b, \alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} y^{(i)} y^{(j)} \alpha_i \alpha_j \boxed{x^{(i)T} x^{(j)}}$$

也可写成 $\phi(x^{(i)})^T \phi x^{(j)}$

③ 对偶转换

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

$$\text{s.t. } \alpha_i\geq 0,\ i=1,\cdots,m$$

由积为 $\phi(x^{(i)})^T\phi(x^{(j)})$

$$\sum_{i=1}^{m}\alpha_i y^{(i)}=0$$

使用kernel替换

$$\max_{\alpha} W(\alpha)=\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{m}y^{(i)}y^{(j)}\alpha_i\alpha_j K(x^{(i)},x^{(j)})$$

$$\text{s.t. } \boxed{\alpha_i\geq 0,\ i=1,\cdots,m}$$

$$\sum_{i=1}^{m}\alpha_i y^{(i)}=0$$

若加入松弛变量，此处为 $0\leq\alpha_i\leq C,\ i=1,\cdots,m$

④ SMO求解 α

以加入松弛变量为例. $\alpha_1 y^{(1)}+\alpha_2 y^{(2)}=-\sum_{i=3}^{m}\alpha_i y^{(i)}=\zeta$

$$\alpha_2=-\frac{y^{(1)}}{y^{(2)}}\alpha_1+\frac{\zeta}{y^{(2)}}$$

$$\alpha_1=(\zeta-\alpha_2 y^{(2)})y^{(1)}$$

$\max_{\alpha} W(\alpha)$ 转化为二次优化问题 $\max_{\alpha_2} W(\alpha_2)=a\alpha_2^2+b\alpha_2+c$

$$\text{s.t. } 0\leq\alpha_2\leq C$$

迭代求解得到所有的 α. 则得到 w, b，但加入Kernel后

$$w=\sum_{i=1}^{m}a_i y^{(i)}\phi x^{(i)},\quad b=-\frac{\max_{i:y^{(i)}=-1}w^T\phi(x^{(i)})+\min_{i:y^{(i)}=1}w^T\phi(x^{(i)})}{2}$$

含有 $\phi(x^{(i)})$. 所以 b 在SMO中求 α 时迭代求得. 预测时

$$\text{predict}=\text{sign}\left(\sum_{i=1}^{n}a_i y^{(i)}K(x^{(i)},x^{new})+b\right)$$

Homework 2

Due: Nov 11, 2020

1. (2 points)

2. (3 points) 评分标准：第一问：结论正确 1 分 第二问：结论正确 1 分，过程 1 分。

(a) We first compute the partial derivatives $\partial y_k / \partial o_i$ of the softmax function

$$y_k = \frac{\exp(o_k)}{\sum_{k'} \exp(o_{k'})}$$

for all i.

- Case 1 ($i = k$):

$$\begin{aligned}\frac{\partial y_k}{\partial o_k} &= \frac{\left(\sum_{k'} \exp(o_{k'})\right) \exp(o_k) - \left(\exp(o_k)\right)^2}{\left(\sum_{k'} \exp(o_{k'})\right)^2} \\ &= y_k - (y_k)^2 \\ &= y_k(1 - y_k).\end{aligned}$$

- Case 2 ($i \neq k$):

$$\begin{aligned}\frac{\partial y_k}{\partial o_i} &= -\frac{\exp(o_k)\exp(o_i)}{\left(\sum_{k'} \exp(o_{k'})\right)^2} \\ &= y_k(0 - y_i).\end{aligned}$$

Combining the two cases, we get

$$\frac{\partial y_k}{\partial o_i} = y_k(\delta_{ki} - y_i). \tag{3}$$

Also, since

$$o_i = \sum_h v_{ih} z_h + v_{i0},$$

we get

$$\frac{\partial o_i}{\partial v_{ih}} = z_h. \tag{4}$$

Likelihood:

$$L(\mathbf{W}, \mathbf{V} \mid \mathcal{X}) = \prod_\ell \prod_k \left(y_k^{(\ell)}\right)^{r_k^{(\ell)}}.$$

Cross-entropy error function:

$$E(\mathbf{W}, \mathbf{V} \mid \mathcal{X}) = -\sum_{\ell}\sum_{k} r_k^{(\ell)} \log y_k^{(\ell)}.$$

Weight update rule for v_{ih}:

$$\begin{aligned}
\Delta v_{ih} = -\eta \frac{\partial E}{\partial v_{ih}} &= \eta \sum_{\ell}\sum_{k} \frac{r_k^{(\ell)}}{y_k^{(\ell)}} \frac{\partial y_k^{(\ell)}}{\partial v_{ih}} \\
&= \eta \sum_{\ell}\sum_{k} \frac{r_k^{(\ell)}}{y_k^{(\ell)}} \frac{\partial y_k^{(\ell)}}{\partial o_i^{(\ell)}} \frac{\partial o_i^{(\ell)}}{\partial v_{ih}} \\
&= \eta \sum_{\ell}\sum_{k} \frac{r_k^{(\ell)}}{y_k^{(\ell)}} y_k^{(\ell)} (\delta_{ki} - y_i^{(\ell)}) z_h^{(\ell)} \qquad \text{(by applying (3) and (4))} \\
&= \eta \sum_{\ell} \left[\sum_{k} r_k^{(\ell)} (\delta_{ki} - y_i^{(\ell)}) \right] z_h^{(\ell)} \\
&= \eta \sum_{\ell} (r_i^{(\ell)} - y_i^{(\ell)}) z_h^{(\ell)}.
\end{aligned}$$

(b) We first compute three derivatives:

$$\begin{aligned}
o_i &= \sum_h v_{ih} z_h + v_{i0} \\
\frac{\partial o_i}{\partial z_h} &= v_{ih} \qquad (5) \\
z_h &= \tanh(a_h) = \frac{e^{a_h} - e^{-a_h}}{e^{a_h} + e^{-a_h}} \\
\frac{d z_h}{d a_h} &= \frac{(e^{a_h} + e^{-a_h})^2 - (e^{a_h} - e^{-a_h})^2}{(e^{a_h} + e^{-a_h})^2} \\
&= 1 - (z_h)^2 \qquad (6) \\
a_h &= \sum_j w_{hj} x_j + w_{h0} \\
\frac{\partial a_h}{\partial w_{hj}} &= x_j. \qquad (7)
\end{aligned}$$

Weight update rule for w_{hj}:

$$\begin{aligned}
\Delta w_{hj} &= -\eta \frac{\partial E}{\partial w_{hj}} \\
&= \eta \sum_{\ell}\sum_{k} \frac{r_k^{(\ell)}}{y_k^{(\ell)}} \frac{\partial y_k^{(\ell)}}{\partial w_{hj}} \\
&= \eta \sum_{\ell}\sum_{k}\sum_{i} \frac{r_k^{(\ell)}}{y_k^{(\ell)}} \frac{\partial y_k^{(\ell)}}{\partial o_i^{(\ell)}} \frac{\partial o_i^{(\ell)}}{\partial z_h^{(\ell)}} \frac{\partial z_h^{(\ell)}}{\partial a_h^{(\ell)}} \frac{\partial a_h^{(\ell)}}{\partial w_{hj}} \\
&= \eta \sum_{\ell}\sum_{k}\sum_{i} \frac{r_k^{(\ell)}}{y_k^{(\ell)}} y_k^{(\ell)} (\delta_{ki} - y_k^{(\ell)}) v_{ih} \big(1 - (z_h^{(\ell)})^2\big) x_j^{(\ell)} \qquad \text{(by applying (3) and (5)–(7))} \\
&= \eta \sum_{\ell} \left[\sum_{i} (r_i^{(\ell)} - y_i^{(\ell)}) v_{ih} \right] \big(1 - (z_h^{(\ell)})^2\big) x_j^{(\ell)}.
\end{aligned}$$