딥러닝 과제05

1 softmax를 적용한 후 출력이 $(0.001, 0.9, 0.001, 0.098)^{T}$이고 레이블 정보가 $(0,0,0,1)^{T}$일 때, 세 가지 목적함수, 평균제곱 오차, 교차 엔트로피, 로그우도를 계산하시오. (로그 우도의 경우 PPT 12쪽에 있는 목적함수를 사용하시오)

$$MSE = \frac{1}{4}\sqrt{(0.001-0)^2+(0.9-0)^2+(0.001-0)^2+(0.098-1)^2} = 0.3186$$

$$CE = -\log(1-0.001) - \log(1-0.9) - \log(1-0.001) - \log(0.098) = 6.6759$$

$$NLL = -\log 0.098 = 3.3511$$

2 [예제 5-1]에서 $\lambda = 0.1, \lambda = 0.5$일 때를 계산하고 λ에 따른 효과를 설명하시오. 이때 [그림 5-21]을 활용하시오.

예제 5-1 리지 회귀

훈련집합 $\mathbb{X} = \{\mathbf{x}_1 = \begin{pmatrix}1\\1\end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix}2\\3\end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix}3\\3\end{pmatrix}\}$, $\mathbb{Y} = \{y_1 = 3.0, y_2 = 7.0, y_3 = 8.8\}$이 주어졌다고 가정하자. 특징 벡터가 2차원이므로 d=2이고 샘플이 3개이므로 n=3이다. 훈련집합으로 설계행렬 **X**와 레이블 행렬 **y**를 다음과 같이 쓸 수 있다.

$$\mathbf{X} = \begin{pmatrix}1 & 1\\2 & 3\\3 & 3\end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix}3.0\\7.0\\8.8\end{pmatrix}$$

이 값들을 식 (5.29)에 대입하여 다음과 같이 $\hat{\mathbf{w}}$을 구할 수 있다. 이때 $\lambda = 0.25$라 가정하자.

$$\hat{\mathbf{w}} = \left(\begin{pmatrix}1 & 2 & 3\\1 & 3 & 3\end{pmatrix}\begin{pmatrix}1 & 1\\2 & 3\\3 & 3\end{pmatrix} + \begin{pmatrix}0.5 & 0\\0 & 0.5\end{pmatrix}\right)^{-1}\begin{pmatrix}1 & 2 & 3\\1 & 3 & 3\end{pmatrix}\begin{pmatrix}3.0\\7.0\\8.8\end{pmatrix} = \begin{pmatrix}1.4916\\1.3607\end{pmatrix}$$

따라서 하이퍼 평면은 $y = 1.4916x_1 + 1.3607x_2$이다. 새로운 샘플로 $\mathbf{x} = (5 \quad 4)^{T}$가 입력되면 식 (5.30)을 이용하여 12.9009를 예측한다.

$$\hat{W} = \left(\begin{pmatrix}1 & 2 & 3\\1 & 3 & 3\end{pmatrix}\begin{pmatrix}1 & 1\\2 & 3\\3 & 3\end{pmatrix} + \begin{pmatrix}0.2 & 0\\0 & 0.2\end{pmatrix}\right)^{-1}\begin{pmatrix}1 & 2 & 3\\1 & 3 & 3\end{pmatrix}\begin{pmatrix}3.0\\7.0\\8.8\end{pmatrix} = \begin{pmatrix}1.6154\\1.2788\end{pmatrix} \Big/ \hat{W} = \left(\begin{pmatrix}1 & 2 & 3\\1 & 3 & 3\end{pmatrix}\begin{pmatrix}1 & 1\\2 & 3\\3 & 3\end{pmatrix} + \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right)^{-1}\begin{pmatrix}1 & 2 & 3\\1 & 3 & 3\end{pmatrix}\begin{pmatrix}3.0\\7.0\\8.8\end{pmatrix} = \begin{pmatrix}1.4\\1.4\end{pmatrix}$$

```
import numpy as np

a = [[1, 2, 3],
     [1, 3, 3]]

b = [[0.2, 0],
     [0, 0.2]]

c = [[3.0],
     [7.0],
     [8.8]]

print(np.matmul(np.matmul(np.linalg.inv(np.matmul(a, np.transpose(a)) + b), a), c))
```

```
[[1.61538462]
 [1.27884615]]
```

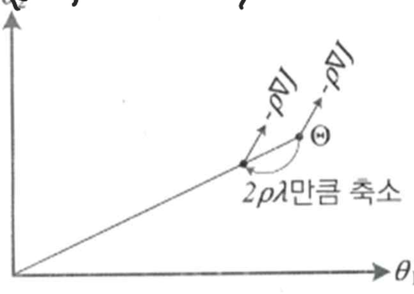

그림 5-21 $L2$ 놈을 사용한 가중치 감쇠 기법의 효과

```
import numpy as np

a = [[1, 2, 3],
     [1, 3, 3]]

b = [[1, 0],
     [0, 1]]

c = [[3.0],
     [7.0],
     [8.8]]

print(np.matmul(np.matmul(np.linalg.inv(np.matmul(a, np.transpose(a)) + b), a), c))
```

```
[[1.4]
 [1.4]]
```

L_2 놈을 사용하여 가중치 감소를 적용하면
Θ를 $2\rho\lambda$의 비율로 줄인 다음 업데이트 하는 것과 동일하다.

3 [알고리즘 5-5]의 Adam은 RMSProp에 식 (5.12)의 모멘텀을 적용하였다. RMSProp에 식 (5.13)의 네스테로프 모멘텀을 적용한 Nesterov_Adam 알고리즘을 작성하시오.

알고리즘 5-5 Adam

입력: 훈련집합 $\mathbb{X}$, $\mathbb{Y}$, 학습률 ρ, 모멘텀 계수 α_1, 가중 이동 평균 계수 α_2

출력: 최적의 매개변수 $\widehat{\boldsymbol{\Theta}}$

난수를 생성하여 초기해 $\boldsymbol{\Theta}$를 설정한다.

$\mathbf{v} = \mathbf{0}, \mathbf{r} = \mathbf{0}$

$t=1$

repeat

~~그레이디언트 $\mathbf{g} = \left.\frac{\partial J}{\partial \boldsymbol{\Theta}}\right|_{\boldsymbol{\Theta}}$ 를 구한다.~~

$\tilde{\theta} = \theta + \alpha_1 v$

$g = \left.\frac{\partial J}{\partial \theta}\right|_{\tilde{\theta}}$

$\mathbf{v} = \alpha_1 \mathbf{v} - (1-\alpha_1)\mathbf{g}$ // 속도 벡터

$\mathbf{v} = \frac{1}{1-(\alpha_1)^t}\mathbf{v}$

$\mathbf{r} = \alpha_2 \mathbf{r} + (1-\alpha_2)\mathbf{g}\odot\mathbf{g}$ // 그레이디언트 누적 벡터

$\mathbf{r} = \frac{1}{1-(\alpha_2)^t}\mathbf{r}$

$\Delta\boldsymbol{\Theta} = -\frac{\rho}{\epsilon+\sqrt{\mathbf{r}}}\mathbf{v}$

$\boldsymbol{\Theta} = \boldsymbol{\Theta} + \Delta\boldsymbol{\Theta}$

t++

until (멈춤 조건)

$\widehat{\boldsymbol{\Theta}} = \boldsymbol{\Theta}$

$$\left.\begin{aligned} \mathrm{v} &= \alpha \mathrm{v} - \rho\frac{\partial J}{\partial \Theta} \\ \Theta &= \Theta + \mathrm{v} \end{aligned}\right\} \cdots\cdots (5.12)$$

$$\left.\begin{aligned} \Theta &= \Theta + \alpha \mathrm{v} \\ \mathrm{v} &= \alpha \mathrm{v} - \rho\left.\frac{\partial J}{\partial \Theta}\right|_{\tilde{\Theta}} \\ \Theta &= \Theta + \mathrm{v} \end{aligned}\right\} \cdots\cdots ($$