4장 연습문제 해답

01.

(a)

$$g(\lambda)=\inf_{x\in\mathbb{R}}L(x,\lambda)=L\left(-\frac{1}{\lambda},\lambda\right)=-\frac{1}{\lambda}+2022+\lambda\cdot\frac{1}{2}\left(-\frac{1}{\lambda}\right)^2=-\frac{1}{2\lambda}+2022$$

(b)

$$\begin{aligned}g(\lambda,\nu)&=\inf_{(x,y)\in\mathbb{R}^2}L(x,y,\lambda,\nu)\\&=L\left(\frac{1}{2}(\lambda-3\nu),\frac{1}{2}(\lambda-2\nu)\right)\\&=\left(\frac{1}{2}(\lambda-3\nu)\right)^2+\left(\frac{1}{2}(\lambda-2\nu)\right)^2+\lambda\left(7-\frac{1}{2}(\lambda-3\nu)-\frac{1}{2}(\lambda-2\nu)\right)+\nu\left(\frac{3}{2}(\lambda-3\nu)+(\lambda-2\nu)-1\right)\\&=-\frac{1}{4}(2\lambda^2-(10\nu-28)\lambda+13\nu^2+4\nu)\end{aligned}$$

02.

$$g(\boldsymbol{\nu})=\begin{cases}-\sum_{i=1}^{n}\nu_i & W+\mathrm{diag}(\boldsymbol{\nu})\geq 0\\ -\infty & otherwise\end{cases}$$

03.

$\mathbb{R}^n_+$에서 정의된 함수 $g(\boldsymbol{x})=\boldsymbol{y}^T\boldsymbol{x}-\sum_{i=1}^{n}x_i\log x_i$의 헤세 행렬을 $\nabla^2 g(\boldsymbol{x})$라 할 때, 헤세 행렬의 정의에 의해 $[\nabla^2 g(\boldsymbol{x})]_{ij}=\frac{\partial^2 g}{\partial x_i \partial x_j}$이므로 $\frac{\partial^2 g}{\partial x_i \partial x_j}$를 구해보면 다음과 같습니다.

$$\frac{\partial^2 g}{\partial x_i\partial x_j}=\frac{\partial}{\partial x_i}\left(\frac{\partial g}{\partial x_j}\right)=\frac{\partial}{\partial x_i}\left(y_j-\left(\log x_j+x_j\cdot\frac{1}{x_j}\right)\right)=\frac{\partial}{\partial x_i}(y_j-(\log x_j+1))=\begin{cases}-\frac{1}{x_i}, & i=j\\ 0\ , & i\neq j\end{cases}$$

따라서 함수 g의 헤세 행렬 $\nabla^2 g(\boldsymbol{x})$은 $[\nabla^2 g(\boldsymbol{x})]_{ii}=-\frac{1}{x_i}$인 대각행렬, 즉

$$\nabla^2 g(\boldsymbol{x})=\begin{bmatrix}-\frac{1}{x_1} & 0 & \cdots & 0\\ 0 & -\frac{1}{x_2} & \cdots & 0\\ \vdots & \cdots & \ddots & \vdots\\ 0 & \cdots & 0 & -\frac{1}{x_n}\end{bmatrix}=-\begin{bmatrix}\frac{1}{x_1} & 0 & \cdots & 0\\ 0 & \frac{1}{x_2} & \cdots & 0\\ \vdots & \cdots & \ddots & \vdots\\ 0 & \cdots & 0 & \frac{1}{x_n}\end{bmatrix}$$

이고, $x_1,\cdots,x_n > 0$이므로 임의의 0이 아닌 $\boldsymbol{a}=(a_1,\cdots,a_n)\in\mathbb{R}^n$에 대해

$$\boldsymbol{a}^T\nabla^2 g(\boldsymbol{x})\boldsymbol{a}=\begin{bmatrix}a_1\ a_2\ \cdots\ a_n\end{bmatrix}\left(-\begin{bmatrix}\frac{1}{x_1} & 0 & \cdots & 0\\ 0 & \frac{1}{x_2} & \cdots & 0\\ \vdots & \cdots & \ddots & \vdots\\ 0 & \cdots & 0 & \frac{1}{x_n}\end{bmatrix}\right)\begin{bmatrix}a_1\\ a_2\\ \vdots\\ a_n\end{bmatrix}=-\frac{1}{x_1}a_1^2-\frac{1}{x_2}a_2^2-\cdots-\frac{1}{x_n}a_n^2<0$$

임을 알 수 있습니다. 따라서, $\nabla^2 g(\boldsymbol{x})$는 음의 정부호 행렬입니다.

04.

(a)

$$\text{maximize} \qquad -pe^{-1-\boldsymbol{\nu}^T\boldsymbol{a}_1}-qe^{-1-\boldsymbol{\nu}^T\boldsymbol{a}_2}-\boldsymbol{\nu}^T\boldsymbol{b}$$

(b)

$$\begin{array}{ll}\text{maximize} & \sum_{i=1}^{m}\log\nu_i+m-\boldsymbol{b}^T\boldsymbol{\nu}\\ \text{subject to} & A^T\boldsymbol{\nu}=\boldsymbol{0}\\ & \nu_i>0,\ i=1,\cdots,m\end{array}$$

05.

$$\begin{array}{ll}\text{maximize} & -\boldsymbol{\lambda}^T\boldsymbol{b}\\ \text{subject to} & \boldsymbol{c}+A^T\boldsymbol{\lambda}=\boldsymbol{0}\\ & \lambda_i\ge 0,\ i=1,\cdots,m\end{array}$$

06.

주어진 볼록 최적화 문제

$$\begin{array}{ll}\text{minimize} & f_0(\boldsymbol{x})\\ \text{subject to} & f_i(\boldsymbol{x})\le 0,\ i=1,\cdots,m\\ & A\boldsymbol{x}=\boldsymbol{b}\end{array}$$

에 대해 집합 A, B를 다음과 같이 정의합시다.

$$A=\{(\boldsymbol{u},\boldsymbol{v},t)|\ \exists\,\boldsymbol{x}\in \mathrm{dom}f,\ f_i(\boldsymbol{x})\le u_i\,(i=1,\cdots,m),\ A\boldsymbol{x}-\boldsymbol{b}=\boldsymbol{0},\ f_0(\boldsymbol{x})\le t\}$$
$$B=\{(\boldsymbol{0},\boldsymbol{0},s)|s<p^*\}$$

여기서 p^*는 위 볼록 최적화 문제의 최적값, 즉 $p^*=f(\boldsymbol{x}^*)$를 의미합니다. 위와 같이 정의된 집합 A, B의 교집합이 없음을 보입시다. 만약, $(\boldsymbol{u},\boldsymbol{v},t)\in A\cap B$라 가정하면, 집합 A의 정의에 의해 어떤 $\boldsymbol{x}$가 존재해서 $f_i(\boldsymbol{x})\le u_i$, $A\boldsymbol{x}-\boldsymbol{b}=\boldsymbol{0}$, $f_0(\boldsymbol{x})\le t$를 만족해야 합니다. 한편, 집합 B의 정의에 의해 $\boldsymbol{u}=\boldsymbol{0}$, $\boldsymbol{v}=\boldsymbol{0}$, $t<p^*$이므로, $f_0(\boldsymbol{x})\le t<p^*$를 만족하는 $\boldsymbol{x}$가 존재하고, 이는 $p^*=f(\boldsymbol{x}^*)$가 볼록 최적화 문제의 최적값이라는 가정에 모순입니다. 따라서, A, B의 교집합이 없습니다.

집합 A와 B의 교집합이 없기 때문에, 집합 A와 B를 구분하는 평면이 존재합니다. 즉, $\boldsymbol{0}$이 아닌 $(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\nu}},\mu)$와 실수 c가 존재해서

$$\begin{aligned}(\boldsymbol{u},\boldsymbol{v},t)\in A \quad&\Rightarrow\quad \hat{\boldsymbol{\lambda}}^T\boldsymbol{u}+\hat{\boldsymbol{\nu}}^T\boldsymbol{v}+\mu t\ \ge c\\ (\boldsymbol{u},\boldsymbol{v},t)\in B \quad&\Rightarrow\quad \hat{\boldsymbol{\lambda}}^T\boldsymbol{u}+\hat{\boldsymbol{\nu}}^T\boldsymbol{v}+\mu t\ \le c\end{aligned}$$

를 만족합니다. 첫 번째 조건($(\boldsymbol{u},\boldsymbol{v},t)\in A\Rightarrow\hat{\boldsymbol{\lambda}}^T\boldsymbol{u}+\hat{\boldsymbol{\nu}}^T\boldsymbol{v}+\mu t\ \ge c$)에서, $\hat{\boldsymbol{\lambda}}^T\boldsymbol{u}+\mu t$가 아래로 유계이기 위해서는 $\hat{\lambda}_i\ \ge 0$, $\mu\ge 0$임을 알 수 있고, 두 번째 조건($(\boldsymbol{u},\boldsymbol{v},t)\in B\Rightarrow$ $\hat{\boldsymbol{\lambda}}^T\boldsymbol{u}+\hat{\boldsymbol{\nu}}^T\boldsymbol{v}+\mu t\ \le c$)은 $t<p^*$인 모든 t에 대해 $\mu t\le c$이므로, $\mu p^*\ \le \alpha$임을 알 수 있습니다. 따라서, 이를 종합하면 $f_i(\boldsymbol{x})<0$, $A\boldsymbol{x}=\boldsymbol{b}$를 만족하는 임의의 $\boldsymbol{x}$에 대해, 다음이 성립합니다.

$$\sum_{i=1}^{m}\hat{\lambda}_i f_i(\boldsymbol{x})+\hat{\boldsymbol{\nu}}^T(A\boldsymbol{x}-\boldsymbol{b})+\mu f_0(\boldsymbol{x})\ge \alpha\ \ge \mu p^*$$

① <u>$\mu>0$**인 경우**</u>

위 부등식의 양변을 μ로 나누면,

$$L\left(\boldsymbol{x},\frac{\hat{\boldsymbol{\lambda}}}{\mu},\frac{\hat{\boldsymbol{\nu}}}{\mu}\right)=\sum_{i=1}^{m}\frac{\hat{\lambda}_i}{\mu}f_i(\boldsymbol{x})+\left(\frac{\hat{\boldsymbol{\nu}}}{\mu}\right)^T(A\boldsymbol{x}-\boldsymbol{b})+f_0(\boldsymbol{x})\ge p^*$$

를 얻습니다. 따라서 $\boldsymbol{\lambda}=\dfrac{\hat{\boldsymbol{\lambda}}}{\mu}$, $\boldsymbol{\nu}=\dfrac{\hat{\boldsymbol{\nu}}}{\mu}$로 두면, $f_i(\boldsymbol{x})<0$, $A\boldsymbol{x}=\boldsymbol{b}$를 만족하는 임의의 $\boldsymbol{x}$에 대해, $g(\boldsymbol{\lambda},\boldsymbol{\nu})\le L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\nu})\ge p^*$를 얻고, 약쌍대성에 의해 $g(\boldsymbol{\lambda},\boldsymbol{\nu})\le p^*$가 성립하므로 $g(\boldsymbol{\lambda},\boldsymbol{\nu})=p^*$, 즉 강쌍대성이 성립합니다.

② $\mu=0$인 경우

이 경우, 위 부등식은

$$\sum_{i=1}^{m}\hat{\lambda}_i f_i(x)+\hat{\nu}^T(Ax-b)\geq 0$$

로 쓸 수 있고, 슬레이터 조건인 $f_i(x)<0$, $Ax=b$를 만족하는 임의의 x에 대해, 이 조건은

$$\sum_{i=1}^{m}\hat{\lambda}_i f_i(x)\geq 0$$

로 쓸 수 있습니다. 한편, $f_i(x)<0$이고 $\hat{\lambda}_i\geq 0$이므로 $\hat{\lambda}_i=0$일때만 이 조건이 성립함을 알 수 있습니다. 따라서, $\hat{\lambda}=0$, $\mu=0$이므로 $(\hat{\lambda},\hat{\nu},\mu)$가 0이 아니라는 가정에 의해 $\hat{\nu}\neq 0$이어야 합니다. 즉, $\sum_{i=1}^{m}\hat{\lambda}_i f_i(x)+\hat{\nu}^T(Ax-b)\geq 0$에 의해 $\hat{\nu}^T(Ax-b)\geq 0$여야 하고, 슬레이터 조건을 만족하는 $\hat{x}$에 대해서 $\hat{\nu}^T(A\tilde{x}-b)=0$이므로, $A^T\hat{\nu}=0$가 아닌 이상 $\hat{\nu}^T(Ax-b)<0$을 만족하는 x가 존재해야 합니다. 이는 A의 랭크가 p라는 가정에 모순입니다.

따라서 ①, ②에 의해 슬레이터 조건을 증명했습니다.

07.

(a) 부등식 제약조건이 없으므로, 슬레이터 조건에 의해 $A^T\begin{bmatrix}x\\y\end{bmatrix}=b$를 만족하는 모든 $\begin{bmatrix}x\\y\end{bmatrix}$에 대해 강쌍대성이 성립합니다.

(b) 부등식 제약조건에서 등호 조건이 없으므로, 슬레티어 조건에 의해 $a_i^Tx<b_i$를 만족하는 모든 x에 대해 강쌍대성이 성립합니다.

08.

주어진 최적화 문제는 볼록 최적화 문제이므로 등식 제약조건에 대한 함수 $h_i(x)$는 어떤 a_i에 대해 $h_i(x)=a_i^Tx$로 표현할 수 있습니다. 따라서, $h_i(x)$의 헤세 행렬은 영행렬이고, x에 대한 라그랑주 함수 $L(x,\lambda,\nu)=f(x)+\sum_{i=1}^{m}\lambda_i f_i(x)+\sum_{i=1}^{p}\nu_i(a_i^Tx)$의 헤세 행렬을 구해보면 다음과 같습니다.

$$\nabla^2 L(x,\lambda,\nu)=\nabla^2 f(x)+\sum_{i=1}^{m}\lambda_i\nabla^2 f_i(x)$$

볼록 최적화 문제의 정의에 의해 목적함수 f와 부등식 제약조건에 대한 함수 f_i 모두 볼록함수이므로 $\nabla^2 f$와 $\nabla^2 f_i$는 양의 준정부호 행렬이고 $\lambda_i \ge 0$이므로, $\mathbf{0}$이 아닌 임의의 $\boldsymbol{x} \in \mathbb{R}^n$에 대해

$$\begin{aligned}\boldsymbol{x}^T \nabla^2 L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \boldsymbol{x} &= \boldsymbol{x}^T \left(\nabla^2 f(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i \nabla^2 f_i(\boldsymbol{x}) \right) \boldsymbol{x} \\ &= \boldsymbol{x}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{x} + \sum_{i=1}^{p} \nu_i \boldsymbol{x}^T \nabla^2 f_i(\boldsymbol{x}) \boldsymbol{x} \ge 0\end{aligned}$$

이 성립합니다. 따라서, 헤세 판정법에 의해 볼록 최적화 문제의 라그랑주 함수는 볼록함수입니다.

09.

주어진 최적화 문제의 쌍대 문제는 다음과 같습니다.

$$\begin{array}{ll} \text{maximize} & \lambda_1 + 4\lambda_2 + 2\lambda_3 \\ \text{subject to} & \begin{array}{l} 1 - \lambda_1 - 2\lambda_2 - \lambda_3 - \lambda_4 = 0 \\ 3 + 2\lambda_1 + 3\lambda_2 - 5\lambda_3 - \lambda_5 = 0 \\ -2 - \lambda_1 + \lambda_2 - \lambda_3 - \lambda_6 = 0 \\ \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6 \ge 0 \end{array} \end{array}$$

구하고자 하는 쌍대 문제의 최적해는 $\boldsymbol{\lambda}^* = \left(\lambda_1^*, \lambda_2^*, \lambda_3^*, \lambda_4^*, \lambda_5^*, \lambda_6^*\right) = \left(\frac{2}{7}, 0, \frac{5}{7}, 0, 0, 1\right)$입니다.

10.

주어진 최적화 문제에서 $f(\boldsymbol{x}) = -\log(\alpha + x) - \log(\beta + y)$, $f_1(\boldsymbol{x}) = -x$, $f_2(\boldsymbol{x}) = -y$, $h(\boldsymbol{x}) = 1 - x - y$로 두면, 주어진 문제에 대한 라그랑주 함수 $L(\boldsymbol{x}, \boldsymbol{\lambda}, \nu)$는 다음과 같습니다.

$$\begin{aligned}L(\boldsymbol{x}, \boldsymbol{\lambda}, \nu) &= f(\boldsymbol{x}) + \sum_{i=1}^{2} \lambda_i f_i(\boldsymbol{x}) + \nu h(\boldsymbol{x}) \\ &= -\log(\alpha + x) - \log(\beta + y) - \lambda_1 x - \lambda_2 y + \nu(1 - x - y)\end{aligned}$$

[정리 4-6]의 KKT 조건을 사용하기 위해 $\nabla_{\boldsymbol{x}} L(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}, \hat{\nu}) = 0$을 만족하는 $\hat{\boldsymbol{x}}$를 계산하면 다음 식을 만족합니다.

$$\begin{aligned}\frac{\partial L}{\partial x} &= -\frac{1}{\alpha + x} - \lambda_1 + \nu = 0 \\ \frac{\partial L}{\partial y} &= -\frac{1}{\beta + y} - \lambda_2 + \nu = 0\end{aligned} \qquad (*)$$

한편, KKT 조건의 조건 ❹에 의해 $\hat{\lambda}_1$, $\hat{\lambda}_2$는 다음을 만족해야 합니다.

$$\hat{\lambda}_1 \hat{x} = 0, \quad \hat{\lambda}_2 \hat{y} = 0$$

(*) 조건과 $\lambda_i \geq 0$을 이용하면,

$$\lambda_1 = \nu - \frac{1}{\alpha + x} \geq 0$$
$$\lambda_2 = \nu - \frac{1}{\beta + y} \geq 0$$

이므로, 상보 여유 조건에 의해

① $\lambda_1 = \nu - \frac{1}{\alpha + x} \neq 0$, 즉, $\nu \geq \frac{1}{\alpha}$이면 $x = 0$,

$\lambda_1 = \nu - \frac{1}{\alpha + x} = 0$이면 $\nu = \frac{1}{\alpha + x}$, 즉 $x = \frac{1}{\nu} - \alpha$

② $\lambda_2 = \nu - \frac{1}{\beta + y} \neq 0$, 즉, $\nu \geq \frac{1}{\beta}$이면 $y = 0$,

$\lambda_2 = \nu - \frac{1}{\alpha + y} = 0$이면 $\nu = \frac{1}{\beta + y}$, 즉 $y = \frac{1}{\nu} - \beta$

임을 알 수 있습니다. 따라서, 제약조건 $x + y = 1$에 의해

$$x + y = 1 \quad \Leftrightarrow \quad \max\left\{0, \frac{1}{\nu} - \alpha\right\} + \max\left\{0, \frac{1}{\nu} - \beta\right\} = 1$$

을 만족하는 ν를 찾음으로써 주어진 문제의 최적해를 찾을 수 있습니다. 이 값은 α와 β에 따라 달라지므로 풀이는 여기서 마치도록 합니다[1].

11.

주어진 최적화 문제의 목적함수를 $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T P \boldsymbol{x} + \boldsymbol{q}^T \boldsymbol{x} + r$이라 하고 등식 제약조건에 대한 함수를 $h(\boldsymbol{x}) = A\boldsymbol{x} - \boldsymbol{b} = 0$이라고 하면, 주어진 최적화 문제의 라그랑주 함수 $L(\boldsymbol{x}, \boldsymbol{\nu})$는 다음과 같습니다.

$$L(\boldsymbol{x}, \boldsymbol{\nu}) = f(\boldsymbol{x}) + \boldsymbol{\nu}^T h(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T P \boldsymbol{x} + \boldsymbol{q}^T \boldsymbol{x} + r + \boldsymbol{\nu}^T (A\boldsymbol{x} - \boldsymbol{b})$$

---

1) 참고로 주어진 최적화 문제는 water-filling이라고 불리는 유명한 최적화 문제입니다. 자세한 내용은 Boyd의 Convex Optimization 책을 참고하세요.

최적해 $\boldsymbol{x}^*$와 ν^*는 KKT 조건에 의해 다음을 만족해야 합니다.

① $\nabla_x L(\boldsymbol{x}^*, \nu^*) = P\boldsymbol{x}^* + q + A^T\nu^* = 0$
② $h(\boldsymbol{x}^*) = A\boldsymbol{x}^* - \boldsymbol{b} = 0$

따라서 이를 행렬 형태로 쓰면

$$\begin{aligned} P\boldsymbol{x}^* + q + A^T\nu^* &= \boldsymbol{0} \\ A\boldsymbol{x}^* &= \boldsymbol{b} \end{aligned} \quad \Leftrightarrow \quad \begin{bmatrix} P & A^T \\ A & O \end{bmatrix} \begin{bmatrix} \boldsymbol{x}^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ \boldsymbol{b} \end{bmatrix}$$

임을 알 수 있습니다.

12.

서포트 벡터 머신을 나타내는 최적화 문제의 목적함수 $f(\boldsymbol{w}) = \frac{1}{2}\|\boldsymbol{w}\|_2^2$는 노름 함수의 제곱이므로 볼록함수이고, 부등식 제약조건에 해당하는 함수 $f_i(\boldsymbol{w}) = 1 - y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \le 0$은 $\boldsymbol{w}$에 대한 선형함수이므로 볼록함수입니다. 따라서, 볼록 최적화 문제의 정의에 의해 서포트 벡터 머신을 나타내는 주어진 최적화 문제는 볼록 최적화 문제입니다.

13.

(a) 여러 가지 답이 존재하지만 이 풀이에서는 직관적인 답을 제시합니다. 서포트 벡터 머신의 라그랑주 함수는 b와 $\boldsymbol{\lambda}$에 대해서는 선형인 함수이고, $\boldsymbol{w}$에 대해서는 전형적인 볼록함수 형태의 이차형식입니다. 따라서, 서포트 벡터 머신의 라그랑주 함수는 볼록함수입니다.

(b) ① $\dfrac{\partial L}{\partial w_j} = w_j - \sum_{i=1}^{N} \lambda_i y_i (\boldsymbol{x}_i)_j$

이차원 데이터 $\boldsymbol{x}_1 = (x_{11}, x_{12})$, $\boldsymbol{x}_2 = (x_{21}, x_{22})$에 대한 서포트 벡터 함수의 라그랑주 함수는 (a)에 의해 다음과 같습니다.

$$L(w_1, w_2, b, \lambda_1, \lambda_2) = \frac{1}{2}(w_1^2 + w_2^2) + \lambda_1[1 - y_1(w_1x_{11} + w_2x_{12} + b)] + \lambda_2[1 - y_2(w_1x_{21} + w_2x_{22} + b)]$$

따라서, 라그랑주 함수를 w_1과 w_2로 각각 미분하면,

$$\frac{\partial L}{\partial w_1} = w_1 + \lambda_1(-y_1x_{11}) + \lambda_2(-y_2x_{21}) = w_1 - \sum_{i=1}^{2} \lambda_i y_i (\boldsymbol{x}_i)_1$$
$$\frac{\partial L}{\partial w_2} = w_2 + \lambda_1(-y_1x_{21}) + \lambda_2(-y_2x_{22}) = w_2 - \sum_{i=1}^{2} \lambda_i y_i (\boldsymbol{x}_i)_2$$

이므로 $\dfrac{\partial L}{\partial w_j} = w_j - \sum_{i=1}^{N} \lambda_i y_i (\boldsymbol{x}_i)_j$ 입니다.

② $\underline{g(\boldsymbol{\lambda}) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j}$

이를 보이기 위해, $\dfrac{\partial L}{\partial b}$를 구해보면 $\dfrac{\partial L}{\partial b} = -\lambda_1 y_1 - \lambda y_2$입니다. 따라서, $g(\boldsymbol{\lambda})$는 $\dfrac{\partial L}{\partial w_i} = 0$, $\dfrac{\partial L}{\partial b} = 0$인 w_i와 b에서의 L의 값을 통해서 얻을 수 있으므로,

$$\begin{aligned} g(\boldsymbol{\lambda}) = g(\lambda_1, \lambda_2) = \inf_{(\mathrm{w}_1, \mathrm{w}_2, \mathrm{b})} & \mathrm{L}(\mathrm{w}_1, \mathrm{w}_2, \mathrm{b}) \\ &= L(\lambda_1 y_1 x_{11} + \lambda_2 y_2 x_{21}, \lambda_1 y_1 x_{21} + \lambda_2 y_2 x_{22}, b) \\ &= \frac{1}{2}\left((\lambda_1 y_1 x_{11} + \lambda_2 y_2 x_{21})^2 + (\lambda_1 y_1 x_{21} + \lambda_2 y_2 x_{22})^2 \right) \\ &\quad + \lambda_1 \left[1 - y_1 \left((\lambda_1 y_1 x_{11} + \lambda_2 y_2 x_{21}) x_{11} + (\lambda_1 y_1 x_{21} + \lambda_2 y_2 x_{22}) x_{12} + b \right) \right] \\ &\quad + \lambda_2 \left[1 - y_1 \left((\lambda_1 y_1 x_{11} + \lambda_2 y_2 x_{21}) x_{21} + (\lambda_1 y_1 x_{21} + \lambda_2 y_2 x_{22}) x_{22} + b \right) \right] \end{aligned}$$

입니다. 위 식을 정리하면 다음을 얻습니다.

$$g(\boldsymbol{\lambda}) = g(\lambda_1, \lambda_2) = \lambda_1 + \lambda_2 - \frac{1}{2}\left(\lambda_1^2 y_1^2 \boldsymbol{x}_1^T \boldsymbol{x}_1 + 2\lambda_1 \lambda_2 y_1 y_2 \boldsymbol{x}_1^T \boldsymbol{x}_2 + \lambda_2^2 y_2^2 \boldsymbol{x}_2^T \boldsymbol{x}_2 \right) = \sum_{i=1}^{2} \lambda_i - \frac{1}{2} \sum_{i=1}^{2} \sum_{j=1}^{2} \lambda_i \lambda_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j$$

14.

(a) ① $\underline{\inf_{\boldsymbol{x}}(f(\boldsymbol{x})) \geq -\sup_{\boldsymbol{x}}(-f(\boldsymbol{x}))}$

$\alpha = \inf_{\boldsymbol{x}}(f(\boldsymbol{x}))$라고 하면, inf의 정의에 의해, 모든 $\boldsymbol{x}$에 대해 $\alpha \leq f(\boldsymbol{x})$이 성립합니다. 이 부등식의 양변에 -1을 곱하면,

$$\alpha \leq f(\boldsymbol{x}) \quad \Leftrightarrow \quad -\alpha \geq -f(\boldsymbol{x})$$

이고, sup의 정의에 의해 $\sup_{\boldsymbol{x}}(-f(\boldsymbol{x})) \geq -\alpha$, 즉

$$-\sup_{\boldsymbol{x}}(-f(\boldsymbol{x})) \leq \alpha = \inf_{\boldsymbol{x}}(f(\boldsymbol{x}))$$

를 얻습니다.

② $\underline{\inf_{\boldsymbol{x}}(f(\boldsymbol{x})) \leq -\sup_{\boldsymbol{x}}(-f(\boldsymbol{x}))}$

$\beta = \sup_{\boldsymbol{x}}(-f(\boldsymbol{x}))$라고 하면 inf의 정의에 의해, 모든 $\boldsymbol{x}$에 대해 $\beta \geq -f(\boldsymbol{x})$이 성립합니다. 이 부등식의 양변에 -1을 곱하면,

$$\beta \geq -f(\boldsymbol{x}) \quad \Leftrightarrow \quad -\beta \leq f(\boldsymbol{x})$$

이고, inf의 정의에 의해 $\inf_{\boldsymbol{x}}(f(\boldsymbol{x})) \leq -\beta$, 즉 다음을 얻습니다.

$$\inf_{\boldsymbol{x}}(f(\boldsymbol{x})) \leq -\beta = -\sup_{\boldsymbol{x}}(-f(\boldsymbol{x}))$$

따라서 ①, ②에 의해 $\inf_{\boldsymbol{x}}(f(\boldsymbol{x})) = -\sup_{\boldsymbol{x}}(-f(\boldsymbol{x}))$이 성립합니다.

(b) $f(x,y) = 5x^2 - 6xy + 5y^2$이 볼록함수임을 보이기 위해, 주어진 함수의 헤세 행렬을 구해보면 다음과 같습니다.

$$[H(f)](x,y) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y} \\ \dfrac{\partial^2 f}{\partial y \partial x} & \dfrac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 10 & -6 \\ -6 & 10 \end{bmatrix}$$

$(0,0)$이 아닌 임의의 (x,y)에 대해,

$$\begin{aligned} [x\ y]\begin{bmatrix} 10 & -6 \\ -6 & 10 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} &= 10x^2 - 12xy + 10y^2 \\ &= 10\left(x^2 - \frac{6}{5}xy + \frac{9}{25}y^2 - \frac{9}{25}y^2\right) + 10y^2 \\ &= 10\left(x - \frac{3}{5}y\right)^2 + \frac{32}{5}y^2 > 0 \end{aligned}$$

임을 알 수 있습니다. 따라서, 헤세 판정법에 의해 주어진 함수는 볼록함수입니다.

15.

[예제 4-11]과 비슷합니다. 직접 해봅시다!