Problem 1 – Maxima of convex functions

1.1 Assume that $x^*$ is not an extreme point. Then there are $x_1, x_2 \in C, \ x_1, x_2 \neq x^*$, so that $x^* = tx_1 + (1 - t)x_2$ for some $t \in (0,1)$. Then,

$$f(x^*) < tf(x_1) + (1 - t)f(x_2) \leq tf(x^*) + (1 - t)f(x^*) = f(x^*)$$

(1)

We have arrived at a contradiction, hence $x^*$ must be an extreme point.

1.2 $x^*$ is not unique. Counterexample: $f(x) = x^2 - 1, \ C = [-1,1]; \ f$ has two maxima at 1 and $-1$.

1.3 $x^*$ is not isolated. Counterexample in $\mathbb{R}^n$: $f(x) = ||x||^2 - 1, \ C = \{||x|| \leq 1\}$. Every point of the boundary of the unit ball is a maximum of $f$, and an extreme point of $C$.

Problem 2 – The rate of convergence of gradient descent with line minimization

2.1

$$g = \nabla f = Qx$$

(2)

$$f(x - \alpha g) = \frac{1}{2}x^TQx + \frac{\alpha^2}{2}g^TQg - \alpha x^TQg$$

(3)

$$\frac{df}{d\alpha}(x - \alpha g) = \alpha g^TQg - x^TQg = 0$$

(4)

$$\alpha = \frac{x^TQg}{g^TQg} = \frac{g^Tg}{g^TQg}$$

(5)

The latter equality follows because $x = Q^{-1}g$.

2.2

$$f(x - \alpha g) = \frac{1}{2} \left[ x^TQx - \frac{(x^TQg)^2}{g^TQg} \right]$$

(6)
\[
\frac{f(x - \alpha g)}{f(x)} = \frac{x^T Q x - \left(\frac{x^T Q g}{g^T Q g}\right)^2}{x^T Q x}
\]
(7)
\[
= 1 - \frac{(x^T Q g)^2}{(g^T Q g) x^T Q x}
\]
(8)
\[
= 1 - \frac{(g^T g)^2}{(g^T Q g)(g^T Q^{-1} g)}
\]
(9)

The latter equality follows because \( x = Q^{-1} g \).

2.3 First, we get the eigenvalues of \( Q \):
\[
\begin{vmatrix}
\lambda - 2 & -a \\
-a & \lambda - 2
\end{vmatrix} = (\lambda - 2)^2 - a^2 = (\lambda - \epsilon)(\lambda - 4 + \epsilon)
\]

It follows that \( \lambda_1 = m = \epsilon, \lambda_2 = M = 4 - \epsilon \). Hence,
\[
\frac{f(x - \alpha g)}{f(x)} \leq 1 - \frac{4mM}{(M + m)^2} = 1 - \frac{4\epsilon(4 - \epsilon)}{4^2} = 1 - \epsilon(1 - \epsilon/4)
\]

For small \( \epsilon \), this rate is nearly 1, and convergence will be very slow.

**Problem 3 – SVM with logarithmic penalty**

\[
\begin{align*}
\min_{w, b, \gamma_{1:m}} & \quad \frac{1}{2} ||w||^2 + \sum_{i=1}^m \ln \frac{1 + e^{\gamma_i}}{2} \\
\text{s.t.} & \quad y_i(w^T x_i + b) \geq 1 - \gamma_i, \text{ for all } i \\
& \quad \gamma_i \geq 0 \text{ for all } i
\end{align*}
\]

(10)

3.1 \( \gamma_{1:m} \) are called *slack variables*. Their role is to measure the amount by which the margin conditions (2) are violated in the solution. \( \gamma_i = 0 \) whenever a point is classified with margin 1 or larger.

3.2, 3.3

\[
L(w, b, \gamma, \lambda, \alpha) = \frac{1}{2} ||w||^2 + \sum_{i=1}^m \ln \frac{1 + e^{\gamma_i}}{2} + \sum_i \lambda_i \left[1 - \gamma_i - y_i(w^T x_i + b)\right] - \sum_i \alpha_i \gamma_i
\]

\[
\frac{\partial L}{\partial w} = w - \sum_i \lambda_i y_i x_i \quad \Rightarrow \quad w = \sum_i \lambda_i y_i x_i
\]
(12)
\[
\frac{\partial L}{\partial b} = \sum_i \lambda_i y_i
\]
(13)
\[
\frac{\partial L}{\partial \gamma_i} = \frac{e^{\gamma_i}}{1 + e^{\gamma_i}} - \lambda_i - \alpha_i
\]
(14)
\[
\gamma_i = -\ln \left(\frac{1}{\lambda_i + \alpha_i} - 1\right)
\]
(15)
It follows that $0 < \alpha_i + \gamma_i < 1$. Denote $K = [K_{ij}]$, $K_{ij} = y_i y_j x_i^T x_j$, $\beta_i = \alpha_i + \lambda_i$. Then

$$1 + e_i^\gamma = \frac{1}{1 - \beta_i}$$

(16)

$$g(\lambda, \alpha) = \frac{1}{2} \lambda^T K \lambda - \sum_i \ln(1 - \beta_i) - m \ln 2 - \sum_i \lambda_i + \sum_i \lambda_i \ln \left( \frac{1}{\beta_i} - 1 \right)$$

$$- \left( \sum_i \lambda_i y_i x_i \right)^T \left( \sum_i \lambda_i y_i x_i \right) - \sum_i \alpha_i \ln \left( \frac{1}{\beta_i} - 1 \right)$$

(17)

$$= -\frac{1}{2} \lambda^T K \lambda - m \ln 2 - \sum_i \lambda_i + \sum_i \left[ - \ln(1 - \beta_i) + \beta_i \ln \frac{1 - \beta_i}{\beta_i} \right]$$

(18)

$$= -\frac{1}{2} \lambda^T K \lambda - m \ln 2 - \sum_i \lambda_i + \sum_i H(\beta_i)$$

(19)

In the above $H(\beta_i)$ denotes the entropy $-\beta_i \ln \beta_i - (1 - \beta_i) \ln(1 - \beta_i)$.

3.4 It is easy to verify that $g$ is concave: the term $-\lambda^T K \lambda$ is a negative quadratic, with $K$ positive definite, the second and third terms are constant, respectively linear, and the terms $H(\beta_i)$ are entropies and therefore concave.

The domain of the dual objective is $\lambda_i \in \mathbb{R}, \beta_i \in (0, 1)$, convex.

$$(D) \max_{\lambda, \alpha} -\frac{1}{2} \lambda^T K \lambda - m \ln 2 - \sum_i \lambda_i + \sum_i H(\beta_i)$$

s.t

$$\lambda_i \geq 0$$

(20)

$$q \beta_i \geq \lambda_i$$

(21)

$$\lambda^T y = 0$$

(22)

All constraints are linear, hence $(D)$ is a concave maximization problem.

3.5 $(D)$ is not a quadratic problem, because of the entropy term $H(\beta_i)$.

3.6 $w^* = \sum_i \lambda_i^* y_i x_i$. Find an $i$ for which $\lambda_i > 0$ Hence, $y_i (w^T x_i + b) = 1 - \gamma_i^* = 1 - \ln \frac{1 - \beta_i^*}{\beta_i^*}$. From this equation, $b^* = y_i^* (1 - \gamma_i^*) - w^T x_i$. The resulting classifier is $f(x) = (x^T w^* + b^*)$.

3.7 If $y_i (x_i^T w + b) < 1$ then $\gamma_i > 0$, then the corresponding $\alpha_i = 0$ by complementary slackness, and $\lambda_i > 0$ because the constraint (10) is tight.

3.8 If $y_i (x_i^T w + b) > 1$ then $\gamma_i = 0$ and the constraint (10) is slack, while the constraint (11) is tight. Hence the corresponding dual variables are $\alpha_i > 0$ and $\lambda_i = 0$, by complementary slackness. For $\gamma_i = 0$ it follows that $\beta_i = \frac{1}{2} = 0 + \alpha_i$, hence $\alpha_i = 1/2$. 

3
3.9 Let \((x_i, y_i)\) be a data point for which \((w^*, b^*)\) has margin = 1. What can you say about \(\lambda_i, \gamma_i, \alpha_i\) in this case? Find \(\lambda_i\) as a function of the other \(\lambda\)'s.

For \(y_i(x_i^T w + b) = 1\) we have \(\gamma_i = 0, \beta_i = 1/2\) as above, and \(\alpha_i > 0\) typically. Hence \(\lambda_i + \alpha_i = 1/2\). The dual objective \(g\) can be written as

\[
g = -\lambda_i^2 K_{ii}/2 - \frac{1}{2} \sum_{j \neq i} \lambda_j K_{ij} \lambda_i + H(\beta_i) + \text{terms independent of } \lambda_i
\]

Also, \(H(\beta_i) = \log 2\) for any \(\lambda_i\). So the maximum over \(\lambda_i\) is attained for

\[
\lambda_i = \begin{cases} 
-k_i / K_{ii} & \text{if } k_i/K_{ii} \in (-1/2, 0) \\
0 & \text{if } k_i \geq 0 \\
\frac{1}{2} & \text{otherwise}
\end{cases}
\]