Machine Learnig Homework 8

Problem 1

The distance to a training sample in feature space is given by

$$|\phi(x) - \phi(x^{(s_i)})|_2. \tag{1}$$

In order that it only depends on the scalar product in feature space $K(x, x^{(s_i)}) = \phi(x)^T\phi(y)$, we square this. Using that operation, the k-nearest neighbours don't change.

$$\begin{aligned} |\phi(x) - \phi(x^{(s_i)})|_2^2 =& (\phi(x) - \phi(x^{(s_i)}))^T(\phi(x) - \phi(x^{(s_i)})) && (2)\\ =& \phi(x)^T\phi(x) - 2\phi(x)^T\phi(x^{(s_i)}) + \phi(x^{(s_i)})^T\phi(x^{(s_i)}). && (3) \end{aligned}$$

When searching for the k training samples that minimize this function, the first term stays constant. Therefore, we can drop the term and must find k training samples $x^{(s_i)}$ that minimize

$$\phi(x^{(s_i)})^T\phi(x^{(s_i)}) - 2\phi(x)^T\phi(x^{(s_i)}) = K(x^{(s_i)}, x^{(s_i)}) - 2K(x, x^{(s_i)}). \tag{4}$$

Problem 2

The definition for a convex function is

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y). \tag{5}$$

Plugging this into $h(x) = f(x) + g(x)$, gives us

$$\begin{aligned} h(\alpha x + (1-\alpha)y) =& f(\alpha x + (1-\alpha)y) + g(\alpha x + (1-\alpha)y) && (6)\\ \leq& \alpha f(x) + (1-\alpha)f(y) + \alpha g(x) + (1-\alpha)g(y) && (7)\\ =& \alpha h(x) + (1-\alpha)h(y). && (8) \end{aligned}$$

The proof of the scaled function being convex is

$$\begin{aligned} u(\alpha x + (1-\alpha)y) =& cf(\alpha x + (1-\alpha)y) && (9)\\ \leq& c\left(\alpha f(x) + (1-\alpha)f(y)\right) && (10)\\ =& \alpha u(x) + (1-\alpha)u(y) && (11) \end{aligned}$$

Problem 3

By definition of convexity for every λ we have for arbitrary x,y and $0 \leq \alpha \leq 1$

$$f_\lambda((1-\alpha)x + \alpha y) \leq (1-\alpha) f_\lambda(x) + \alpha f_\lambda(y). \tag{12}$$

By definition of maximum we have

$$(1-\alpha)\max_\lambda f_\lambda(x) + \alpha \max_\lambda f_\lambda(y) \geq (1-\alpha) f_{\lambda'}(x) + \alpha f_{\lambda'}(y) \tag{13}$$

for any λ', which is the maximizer of $f_\lambda((1-\alpha)x + \alpha y)$ w.r.t. λ. We obtain

$$\begin{aligned}(1-\alpha) f_{\lambda'}(x) + \alpha f_{\lambda'}(y) \geq & f_{\lambda'}(1-\alpha)x + \alpha y) & (14)\\ = & \max_\lambda f_\lambda((1-\alpha)(x + \alpha y). & (15)\end{aligned}$$

s

$$g(\alpha) = \min_x L_x(\alpha) = -\max_x(-L_x(\alpha)). \tag{16}$$

The negative Lagragian $-L_x(\alpha)$ is also convex, because the negation only changes the sign of the constants $C_{0,x}$ and $C_{i,x}$. Thus, as shown in the previous exercise,$-g(\alpha)$ is convex and $g(\alpha)$ is concave

Problem 5

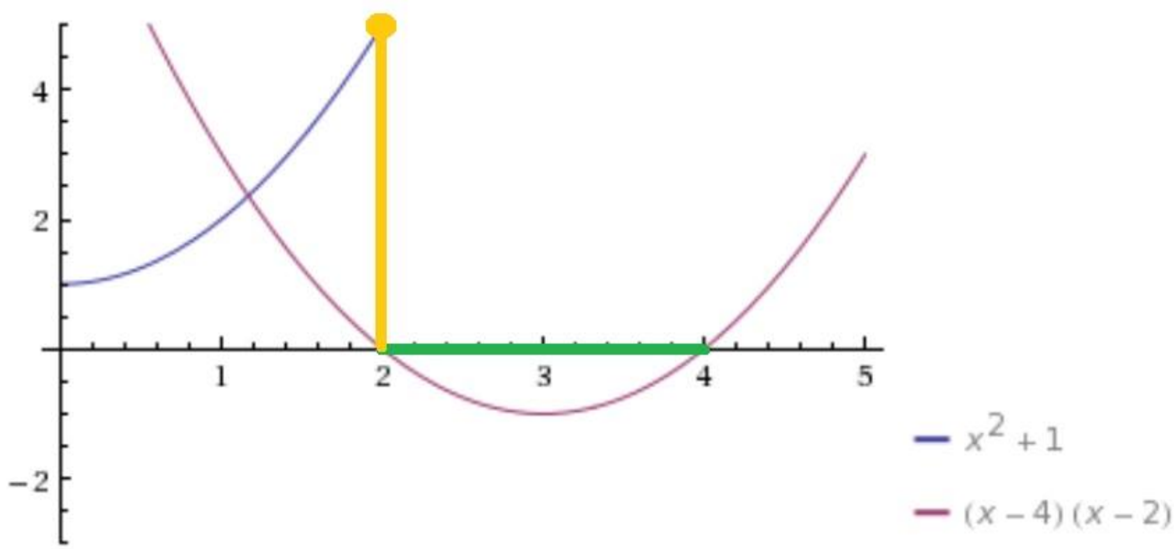

Figure 1: The objective $f_0(x)$ and its constraint $f_1(x)$.

The feasible region is marked green which is from 2 to 4 and the solution of the optimization problem is 2, also called the minimizer. In figure 1 the minimizer is marked orange.

Problem 6

$$\begin{aligned} L(x, \alpha) &= f_0(x) + \alpha f_1(x) \\ &= x^2 + 1 + \alpha(x-2)(x-4) \\ &= (1+\alpha)x^2 - 6\alpha x + (1 + 8\alpha) \end{aligned} \tag{17}$$

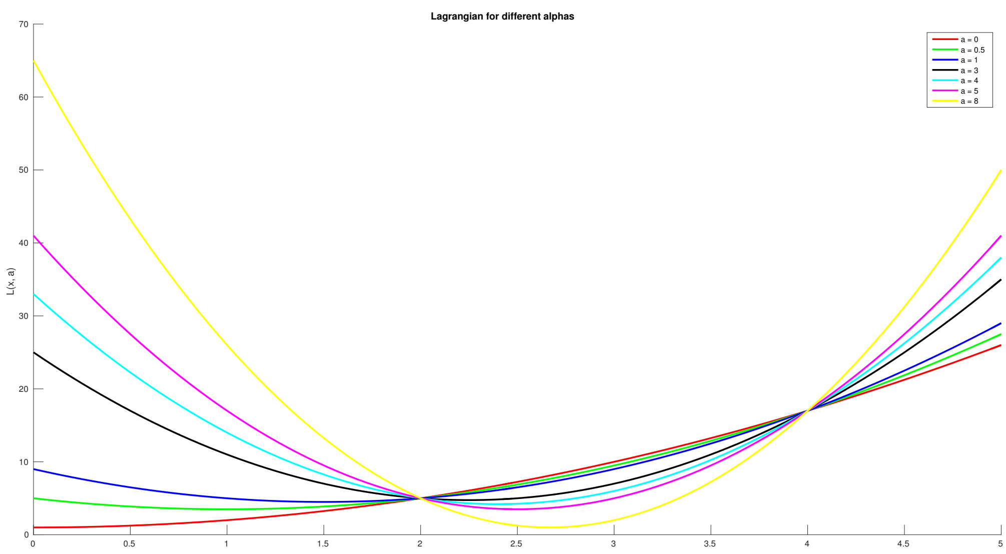

Figure 2: Lagrangian $L(x, \alpha)$ for different values of α

.

The value of the objective function is given by the lagrangian $L(x, \alpha)$ evaluated at $\alpha = 0$, portrayed by the red curve in Fig. 2. Lagrangian is smaller than the objective function for $x \in (2, 4)$ and greater than the objective function for $x \notin [2, 4]$. Values at $x = 2$ and $x = 4$ are unaffected by α. The upper bound for $\min_x L(x, \alpha) = L(2, \alpha)$ for all $\alpha \geq 0$ is 5.

Problem 7

$$g(\alpha) = \min_x L(x, \alpha) \implies \frac{\partial}{\partial x} L(x, \alpha) = 0 \implies x^\star = \frac{3\alpha}{1 + \alpha} \tag{18}$$

$$g(\alpha) = \frac{-\alpha^2 + 9\alpha + 1}{1 + \alpha} \tag{19}$$

The dual problem is plotted in Fig. 3 and it is given by

$$
\begin{aligned}
&\max_{\alpha} g(\alpha) \\
&\text{subject to } \alpha \geq 0
\end{aligned} \tag{20}
$$

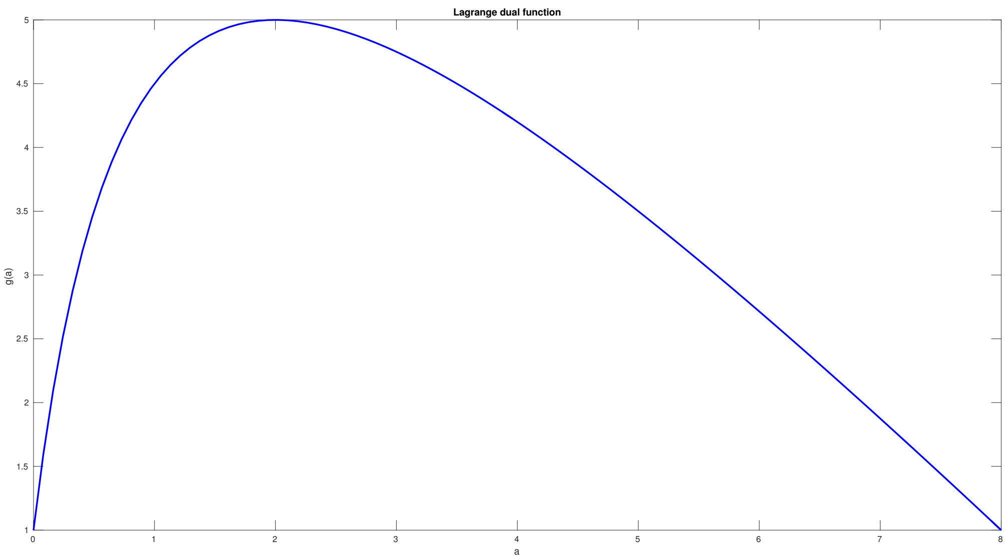

Figure 3: Lagrange dual function $g(\alpha)$.

Problem 8

$$
\alpha^\star = \arg\max_{\alpha} g(\alpha) \tag{21}
$$

$$
0 = \frac{\partial}{\partial \alpha} g(\alpha) = \frac{-\alpha^2 - 2\alpha + 8}{(1+\alpha)^2} \text{ and } \alpha \geq 0 \tag{22}
$$

The dual optimal solution $\alpha^\star = 2$ and $g(\alpha^\star) = 5$.

Problem 9

$$
x^\star = \frac{3\alpha^\star}{1+\alpha^\star} = \frac{6}{1+2} = 2 \tag{23}
$$

The optimal solution is given by $f_0(x^\star) = 5$, which is equal to the dual optimal value.

Problem 10

The constraint f_1 is active. We can see it also on the plot of the primal problem in Fig. 1, since the feasible region is bounded by the zero-crossings of this constraint and the objective function attains its minimum at one of them.