University of Salzburg | Lecturer: Roland Kwitt

Machine Learning (911.236)

Exercise sheet **E**

Below are few relevant lemmas and definitions (repeated from the book) required to solve the exercises. Knowledge of what a convex function is, is assumed.

Lemma 1. *Let $f : \mathbb{R} \to \mathbb{R}$ be a twice differentiable function and let f', f'' denote its first and second derivative, respectively. Then, the following statements are equivalent:*

1. *f is convex*
2. *f' is monotonically nondecreasing*
3. *f'' is nonnegative*

Lemma 2. *Assume that $f : \mathbb{R}^d \to \mathbb{R}$ can be written as*

$$f(\mathbf{w}) = g(\langle \mathbf{w}, \mathbf{x} \rangle + b)$$

for $\mathbf{x} \in \mathbb{R}^d, \mathbf{w} \in \mathbb{R}^d, b \in \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$. Then convexity of g implies convexity of f.

Definition 1. *Let $C \subset \mathbb{R}^d$. A function $\mathbb{R}^d \to \mathbb{R}^k$ is α-Lipschitz over C, if for every $\mathbf{u}, \mathbf{v} \in C$ we have that*

$$\|f(\mathbf{u}) - f(\mathbf{v})\| \leq \alpha \|\mathbf{u} - \mathbf{v}\| \enspace .$$

Note that if $f : \mathbb{R} \to \mathbb{R}$ is differentiable, then, by the mean value theorem (see here), we have

$$f(u) - f(v) = f'(w)(u - v) \enspace ,$$

where w is some point between u and v. It follows that if the derivative of f is everywhere bounded (in absolute value) by α, then the function f is α-Lipschitz.

Lemma 3. *Let $f(\mathbf{x}) = g_1(g_2(\mathbf{x}))$ with g_1 α-Lipschitz and g_2 β-Lipschitz. Then, f is $(\alpha\beta)$-Lipschitz. In particular, if*

$$g_2(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b$$

for some $\mathbf{w} \in \mathbb{R}^d$ and $b \in \mathbb{R}$, then f is $(\alpha\|\mathbf{w}\|)$-Lipschitz.

Exercise 1. 3 P.

Show that the loss function used in logistic regression, i.e.,

$$l(\mathbf{w}, (\mathbf{x}, y)) = \log\left(1 + \exp(-y\langle \mathbf{w}, \mathbf{x} \rangle)\right), \quad \mathbf{w} \in \mathbb{R}^d, \mathbf{x} \in \mathbb{R}^d, y \in \{+1, -1\}$$

is *convex* (in its first argument, i.e., $l(\cdot, (x, y))$ using Lemmas 1 and 2.

Exercise 2. 3 P.

Consider the hypothesis class

$$\mathcal{H}_{\text{sig}} = \{\mathbf{x} \mapsto \phi_{\text{sig}}(\langle \mathbf{w}, \mathbf{x} \rangle) : \mathbf{w} \in \mathbb{R}^d, \|\mathbf{w}\| \leq B\} \enspace ,$$

where ϕ_{sig} is the sigmoid function. This is the hypothesis class we use in logistic regression (with a bound on the norm of the $\mathbf{w}$ here). Show that the loss function

$$l(\mathbf{w}, (\mathbf{x}, y)) = \log\left(1 + \exp(-y\langle \mathbf{w}, \mathbf{x} \rangle)\right), \quad \mathbf{w} \in \mathbb{R}^d, \mathbf{x} \in \mathbb{R}^d, y \in \{+1, -1\}$$

is B-*Lipschitz* continuous (see Def. 1). Strategy: First, show that $f(x) = \log(1 + \exp(x))$ is 1-Lipschitz and then use Lemma 3.

Exercise 3. 2 P.

Let $\mathcal{X} = \mathbb{R}^d$ and $\mathcal{Y} \in \{+1, -1\}$. Show that the *decision boundary* of a logistic regression classifier is, in fact, a hyperplane. Hint: Think about the case where the loss for $y = -1$ is the same as for $y = +1$.

Exercise 4. 2 P.

Consider the XOR function with domain $\mathcal{X} = \{0, 1\} \times \{0, 1\}$ and truth table given below:

	x_1	x_2	$\text{XOR}(x_1, x_2)$
$\mathbf{x}_1$	0	0	0
$\mathbf{x}_2$	0	1	1
$\mathbf{x}_3$	1	0	1
$\mathbf{x}_4$	1	1	0

Table 1: XOR truth table.

Lets abuse the definition of the XOR function a little and think of the output 0 as -1. First, argue why these four points are *not linearly separable*. Second, can you find a strategy to make this data linearly separable (using only x_1 and x_2)? If so, provide the parameters of a separating *halfspace hypothesis*. Hint: Feel free to add dimensions :)

Total #points: 10 P.