Remark: Tools needed for solving Assignment 2 and 3 are discussed during the lecture on SVMs which will take place on Tuesday 23, 2018.

Assignment 1. Let \( \{(x^i, y^i) \in (\mathcal{X} \times \{+1, -1\}) \mid i = 1, \ldots, m\} \) be a training set containing \( m \) examples, \( \phi: \mathcal{X} \to \mathbb{R}^n \) be a fixed feature map and \( \lambda > 0 \) a given scalar. The linear SVM algorithm translates learning of parameters \((w, b) \in \mathbb{R}^{n+1}\) of a linear classifier \( h(x; w, b) = \text{sign}(\langle \phi(x), w \rangle + b) \) into the following convex unconstrained problem

\[
(w_\lambda, b_\lambda) = \arg \min_{(w, b) \in \mathbb{R}^{n+1}} \left( \frac{\lambda}{2} \|w\|^2 + \frac{1}{m} \sum_{i=1}^{m} \max\{0, 1 - y^i(\langle w, \phi(x^i) \rangle + b)\} \right).
\] (1)

An alternative reformulation of the problem (1) reads

\[
(w_R, b_R) = \arg \min_{\|w\| \leq R, b \in \mathbb{R}} \left( \frac{1}{m} \sum_{i=1}^{m} \max\{0, 1 - y^i(\langle w, \phi(x^i) \rangle + b)\} \right).
\] (2)

Prove that for any \( \lambda > 0 \) there exists \( R \geq 0 \) such that \((w_R, b_R)\) obtained as a solution of (2) is also a solution of (1).

Assignment 2. Assume we are given a training set of examples \( T^m = \{(x^i, y^i) \in (\mathcal{X} \times \{+1, -1\}) \mid i = 1, \ldots, m\} \) which is known to be linearly separable with respect to a feature map \( \phi: \mathcal{X} \to \mathbb{R}^n \). In this case, we can find parameters \((w, b) \in \mathbb{R}^{n+1}\) of a linear classifier \( h(x; w, b) = \text{sign}(\langle \phi(x), w \rangle + b) \) which has zero training error by the Perceptron algorithm:

1. \( w \leftarrow 0, b \leftarrow 0 \)

2. Find an example \((x^u, y^u) \in T^m\) whose label is incorrectly predicted by the current classifier, that is \( h(x^u; w, b) \neq y^u \).

3. If all examples are classified correctly exit the algorithm. Otherwise update the parameters by

\[
(w \leftarrow w + y^u \phi(x^u) \quad \text{and} \quad b \leftarrow b + y^u
\]

and go to Step 2.

Assume that you cannot evaluate the feature map \( \phi(x) \) because it is either unknown or its evaluation is expensive. However, you know how to cheaply evaluate a kernel function \( k: \mathcal{X} \times \mathcal{X} \to \mathbb{R} \) such that \( k(x, x') = \langle \phi(x), \phi(x') \rangle \), \( \forall x, x' \in \mathcal{X} \). Show that you can still use the Perceptron algorithm to find a linear classifier with zero training error and that you can evaluate this classifier on any \( x \in \mathcal{X} \).
Assignment 3. Let the input observation be a vector $x \in \mathbb{R}^d$. Let us consider a feature map $\phi_q : \mathbb{R}^d \to \mathbb{R}^n$, $n = d^q$, whose entries are all possible $q$-th degree ordered products of the entries of $x$. For example, if $x = (x_1, x_2, x_3)^T \in \mathbb{R}^3$ and $q = 2$ then

$$
\phi_q(x) = \begin{pmatrix}
x_1 x_1 \\
x_2 x_1 \\
x_3 x_1 \\
x_1 x_2 \\
x_2 x_2 \\
x_3 x_2 \\
x_1 x_3 \\
x_2 x_3 \\
x_3 x_3 
\end{pmatrix}
$$

a) Show that for any $x, x' \in \mathbb{R}^d$ we can compute the dot product between $\phi_q(x)$ and $\phi_q(x')$ as

$$
\langle \phi_q(x), \phi_q(x') \rangle = \langle x, x' \rangle^q,
$$

that is, as the dot product of the original vectors $x$ and $x'$ powered to $q$.

b) Consider a slightly different feature map $\phi' : \mathbb{R}^d \to \mathbb{R}^d(d+1)/2$ whose entries are

$$
\phi'(x) = \begin{pmatrix}
x_1^2 \\
x_2^2 \\
\sqrt{2} x_1 x_2 \\
\sqrt{2} x_1 x_3 \\
\vdots \\
x_d^2 
\end{pmatrix}^T,
$$

so that the features correspond to all possible products of unordered pairs of entries from $x$, and the products of different entries are multiplied by a constant factor $\sqrt{2}$. For example, if $x = (x_1, x_2, x_3)^T \in \mathbb{R}^3$ then

$$
\phi'(x) = (x_1^2, \sqrt{2} x_1 x_2, \sqrt{2} x_1 x_3, \ldots, \sqrt{2} x_1 x_d, x_2^2, \sqrt{2} x_2 x_3, \ldots, \sqrt{2} x_2 x_d, \ldots, x_d^2)^T.
$$

This feature map defines a kernel $k(x, x') = \langle \phi'(x), \phi'(x') \rangle$ referred to as the homogeneous polynomial kernel of degree 2. Show that the kernel value equals to the square of the dot product of the input vectors, that is prove the identity

$$
k(x, x') = \langle \phi'(x), \phi'(x') \rangle = \langle x, x' \rangle^2, \quad \forall x, x' \in \mathbb{R}^d.
$$

**Hint:** Exploit the relation between $\phi(x)$ and $\phi'(x)$.

Assignment 4. Consider a hypothesis space of classifiers

$$
\mathcal{H} = \{ h(x; a) = \text{sign}(\sin(ax)) \mid a \in \mathbb{R} \}.
$$

That is, each $h \in \mathcal{H}$ is determined by a single parameter $a \in \mathbb{R}$ and it maps real valued input $x \in \mathbb{R}$ to a set of hidden labels $\{+1, -1\}$ based on the sign of the score $\sin(ax)$. Show that the VC dimension of $\mathcal{H}$ is infinite.
Hint: Show that for arbitrary set of labels \( \{y^i \in \{+1, -1\} \mid i = 1, \ldots, m\} \) the inputs \( \{x^i = 10^{-i} \mid i = 1, \ldots, m\} \) can be predicted correctly by \( h(x; a) \) with

\[
a = \pi \left( 1 + \frac{1}{2} \sum_{i=1}^{m} (1 - y^i)10^i \right)
\]

Assignment 5. Let us consider the space of all linear classifiers mapping \( x \in \mathbb{R}^d \) to \( \{-1, +1\} \), that is

\[
H = \{ h(x; w, b) = \text{sign}(\langle w, x \rangle + b) \mid (w, b) \in (\mathbb{R}^d \times \mathbb{R}) \}.
\]

Show that the VC dimension of \( H \) is \( d + 1 \).