Homework
Due January 21st

1 Logistic discriminant classification for two classes

The logistic discriminant classifier is given by:

$$p(y = +1|\mathbf{x}) = \sigma(\mathbf{w}^T\mathbf{x}),$$

where $y \in \{-1, +1\}, \mathbf{x} \in \mathbb{R}^p, \mathbf{w} \in \mathbb{R}^p$, and the sigmoid function is given by

$$\sigma(z) = (1 + \exp(-z))^{-1}.$$

The logistic loss for a training sample $\mathbf{x}_i$ with class label y_i is given by:

$$L(y_i, \mathbf{w}^T\mathbf{x}_i) = -\log p(y_i|\mathbf{x}_i).$$

1. Show that $p(y = -1|\mathbf{x}) = \sigma(-\mathbf{w}^\top\mathbf{x})$.
2. Derive that the gradient of the logistic loss has the form
$$\nabla_w L(y_i, \mathbf{w}^\top\mathbf{x}_i) = -y_i\Big(1 - p(y_i|\mathbf{x}_i)\Big)\mathbf{x}_i.$$
3. Show that the logistic loss function is convex in $\mathbf{w} \in \mathbb{R}^p$.

2 Combination rules for kernels

Consider a set $\mathcal{X}$ and two positive definite (p.d.) kernels $K_1, K_2 : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$.

1. For all scalars $\alpha, \beta \geq 0$, show that the sum kernel $\alpha K_1 + \beta K_2$ is p.d.
2. Show that the product kernel $(x, y) \mapsto K_1(x, y)K_2(x, y)$ is p.d. (Be careful, this is a pointwise multiplication, not a matrix multiplication)
    - Tip1: It is sufficient to show that for any finite set $x_1, \ldots, x_n$ in $\mathcal{X}$, the corresponding $n \times n$ kernel matrix is positive semi-definite. Remember that a positive semi-definite matrix A in $\mathbb{R}^{n\times n}$ can be factorized into a product $Z^\top Z$ with Z in $\mathbb{R}^{n\times n}$.
    - Tip2: remember the proof for showing that the polynomial kernel $(x, y) \mapsto (x^\top y)^2$ is p.d. with the "trace" trick. This was indeed a product kernel with $K_1(x, y) = K_2(x, y) = (x^\top y)$ and $\mathcal{X} = \mathbb{R}^p$.

3. Given a sequence $(K_n)_{n\geq 0}$ of p.d. kernels such that for all x, y in $\mathcal{X}$, $K_n(x, y)$ converges to a value $K(x, y)$ in $\mathbb{R}$ (pointwise convergence). Show that K is a p.d. kernel.

4. Show that e^{K_1} is p.d.

3 Positive definite kernels

Which of these kernels are positive definite. You need to provide proofs for all cases.

- $K(x, y) = 1/(1 - xy)$ with $\mathcal{X} = (-1, 1)$.
- $K(x, y) = \min(x, y)$ with $\mathcal{X} = \mathbb{N}$.
- $K(x, y) = \max(x, y)$ with $\mathcal{X} = \mathbb{N}$.
- $K(x, y) = \cos(x + y)$ with $\mathcal{X} = \mathbb{R}$.
- $K(x, y) = \cos(x - y)$ with $\mathcal{X} = \mathbb{R}$.
- $K(x, y) = GCD(x, y)$ (greatest common divisor) with $\mathcal{X} = \mathbb{N}$.

4 The kernel $K(x, y) = \min(x, y)$ with $\mathcal{X} = [0, 1]$

A function $f : \mathcal{X} \to \mathbb{R}$ is said to be "absolutely continuous" if the function is differentiable almost everywhere with f' integrable and $f(x) = f(0) + \int_{t=0}^{x} f'(t)dt$ for all x in $\mathcal{X}$. Surprisingly, we will see that this concept is closely related to the min kernel.

1. Show that the kernel K is p.d.
   Tip: try to write $\min(x, y)$ as an integral.

2. Show that the functional space $\mathcal{H}$ below is Hilbertian (you need to define an appropriate inner-product):

   $$\mathcal{H} = \left\{ f : \mathcal{X} \to \mathbb{R} \text{ such that } f \text{ is absolutely continuous, } f' \in L^2(\mathcal{X}) \text{ and } f(0) = 0 \right\},$$

   where

   $$L^2(\mathcal{X}) = \left\{ f : \mathcal{X} \to \mathbb{R} \text{ such that } \int_{t\in\mathcal{X}} f(t)^2 dt < +\infty \right\}$$

   is a Hilbert space with inner product $\langle f, g\rangle_{L^2(\mathcal{X})} = \int_{t\in\mathcal{X}} f(t)g(t)dt$. [1]

3. Show that K is the reproducing kernel associated to $\mathcal{H}$ for an appropriate inner-product.

Remark: even though, we do not precise it in the text for simplicity, we always consider Lebesgue measurable functions.

[1] Note that to be rigorous, we should remark that $L^2(\mathcal{X})$ is a Hilbert space only when two functions that are equal almost everywhere are considered to be the same element of $L^2(\mathcal{X})$. This can be formalized by defining a quotient space with an equivalence class, or by using the concept of distribution. To simplify, we will omit this subtlety, since it will not affect any result of the homework.