Machine Learning	Dr. Mario Fritz and Dr. Björn Andres
Exercise Sheet 3 - 14.11.2014	due: 21.11.2014

Exercise 9 - Bayes Classifier

Consider a binary classification problem, i.e. the random variable Y takes values in $\{-1, 1\}$ and we have a d-dimensional feature vector X. The class-conditional distributions are given as d-dimensional Gaussians with mean vectors μ_+ and μ_- and equal covariance matrix Σ, i.e.

$$\begin{aligned} p(x|Y=1) &= \frac{1}{(2\pi)^{d/2}\det(\Sigma)^{1/2}}\, e^{-\frac{1}{2}\left\langle x-\mu_+, \Sigma^{-1}(x-\mu_+)\right\rangle}, \\ p(x|Y=-1) &= \frac{1}{(2\pi)^{d/2}\det(\Sigma)^{1/2}}\, e^{-\frac{1}{2}\left\langle x-\mu_-, \Sigma^{-1}(x-\mu_-)\right\rangle} \end{aligned}$$

a. **(2 Points)** Verify that the decision boundary, this is the set $\{x \in \mathbb{R}^d \,|\, \mathrm{P}(Y=1 \,|\, X=x) = \mathrm{P}(Y=-1 \,|\, X=x)\}$, is the hyperplane given as

$$\left\{ x \in \mathbb{R}^d :\ \left\langle x, \Sigma^{-1}(\mu_+ - \mu_-)\right\rangle = \frac{\left\langle \mu_+, \Sigma^{-1}\mu_+\right\rangle - \left\langle \mu_-, \Sigma^{-1}\mu_-\right\rangle}{2} + \log\left(\frac{\mathrm{P}(Y=-1)}{\mathrm{P}(Y=1)}\right)\right\}.$$

b. **(3 Points)** We now specialize to the case $d = 1$ in which case $\Sigma = \sigma^2$ is a positive scalar. Moreover, assume that $\mathrm{P}(Y=1) = \mathrm{P}(Y=-1) = \frac{1}{2}$. Show that the Bayes error R^* of this classification problem is given by

$$R^* = \mathrm{P}\left(U > \frac{|\mu_+ - \mu_-|}{2\sigma}\right), \text{ where } U \sim \mathcal{N}(0,1).$$

Hints:

- You first have to derive $\mathrm{P}(Y=1 \,|\, X=x)$ resp. $\mathrm{P}(Y=-1 \,|\, X=x)$ using Bayes law.
- In order to establish b., it is helpful to note that if $X \sim \mathcal{N}(\mu, \sigma^2)$, then $(X-\mu)/\sigma \sim \mathcal{N}(0,1)$.

Exercise 10 - ML and MAP Estimation

We have as likelihood function,

$$p(x \,|\, \theta) = \begin{cases} \theta\, e^{-\theta x}, & x \geq 0, \\ 0, & x < 0. \end{cases},$$

which is the exponential density with parameter θ.

a. **(2 Points)** Derive the maximum-likelihood estimate of θ given n samples $x_1, \ldots, x_n$.

b. **(2 Points)** We have now an additional prior for the value of θ given as

$$p(\theta) = \begin{cases} \gamma\, e^{-\gamma\theta}, & \theta \geq 0, \\ 0, & \theta < 0. \end{cases},$$

for some fixed value of γ. Derive the MAP estimator of θ.

Hints:

- You may assume that $-\log(p(x \,|\, \theta))$ is convex in θ.

Exercise 11 - Projected Gradient Descent for Lasso

Let $y \in \mathbb{R}^n$ be the n outputs and $\Phi \in \mathbb{R}^{n \times D}$ the design matrix of a regression problem (D basis functions $\phi_1, \ldots, \phi_D$, then $\Phi_{ij} = \phi_j(x_i)$, $i = 1, \ldots, n$, $j = 1, \ldots, D$).

In the lecture the Lasso problem

$$\min_{w \in \mathbb{R}^D} \frac{1}{n} \|Y - \Phi w\|_2^2 + \lambda \|w\|_1 \tag{1}$$

has been rewritten into the optimization problem

$$\begin{aligned} \min_{w^+, w^- \in \mathbb{R}^D} \quad & \frac{1}{n} \left\| Y - \Phi w^+ + \Phi w^- \right\|_2^2 + \lambda \sum_{i=1}^{D} w_i^+ + \lambda \sum_{i=1}^{D} w_i^- \qquad (2) \\ \text{subject to:} \quad & w_i^+ \geq 0, \; i = 1, \ldots, D, \\ & w_i^- \geq 0, \; i = 1, \ldots, D. \end{aligned}$$

Given projection $P_C : \mathbb{R}^d \to \mathbb{R}$ onto a convex set C is defined for $x \in \mathbb{R}^d$ as

$$P_C(x) := \arg\min_{y \in C} \frac{1}{2} \|x - y\|_2^2 .$$

a. **(2 points)** Show that the projection onto a convex set is uniquely defined (Under which condition on the objective is the global minimum unique ?)

b. **(2 points)** Derive an analytical expression for the projection onto the convex set

$$C = \{x \in \mathbb{R}^d \,|\, x_i \geq 0\}, \quad \text{(positive orthant in } \mathbb{R}^d).$$

Hints:

- A sum $f + g$ of convex functions f, g is strictly convex if f or g is strictly convex.
- For the computation of the projection, note that each component can be minimized independently of the other ones. Why ?