COMP4680/COMP8650: Advanced Topics in Machine Learning

Assignment #4: Applications

Due: 11:55pm on Sunday 2 October, 2022.
Submit solutions (showing all working) as a single PDF file or Jupyter notebook via Wattle. Don't forget to include your name and student ID at the top of the submitted solutions. Attribute/cite all sources.

1. **Conjugate functions (20 marks).** Recall the definition of a conjugate function is $f^*(y) = \sup_{x \in \mathbf{dom}(f)} \left\{ x^T y - f(x) \right\}$.
   (a) Show $\inf_x f(x) = -f^*(0)$.
   (b) Show $f(x) + f^*(y) \geq x^T y$ for all x, y.
   (c) Show that $f^{**}(x) \leq f(x)$ for all x. Here $= \sup_{y \in \mathbf{dom}(f^*)} \{x^T y - f^*(y)\}$.
   (d) Compute f^* for $f(x) = \|x\|$ for arbitrary norm $\| \cdot \|$.

2. **Norm approximation and least norm (20 marks).** What is the solution of the following norm approximation and least norm problems with one scalar variable $x \in \mathbb{R}$,
   (a) minimize $\|x\mathbf{1} - a\|_1$
   (b) minimize $\|x\mathbf{1} - a\|_2$
   (c) minimize $\|x\mathbf{1} - a\|_\infty$
   (d) minimize $\|x\|_p$ subject to $a + x \preceq \mathbf{1}$

   for $a \in \mathbb{R}^n$ and $x \in \mathbb{R}$? Here $\mathbf{1} \in \mathbb{R}^n$ is the all-ones vector.

3. **Dual penalty function approximation problems (30 marks).** Derive the Lagrange dual for the problem

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m \phi(r_i) \\ \text{subject to} & r = Ax - b, \end{array}$$

   for the following penalty function $\phi : \mathbb{R} \to \mathbb{R}$. The variables are $x \in \mathbb{R}^n$ and $r \in \mathbb{R}^m$.

   (a) *Deadzone-linear penalty* (with deadzone width $\alpha = 1$),

$$\phi(u) = \begin{cases} 0 & |u| \leq 1, \\ |u| - 1 & |u| > 1. \end{cases}$$

   (b) *Scaled Huber penalty* (with $M = 1$),

$$\phi(u) = \begin{cases} u^2 & |u| \leq 1, \\ 2|u| - 1 & |u| > 1. \end{cases}$$

   (c) *Lopsided linear penalty*

$$\phi(u) = \begin{cases} 2u & u \geq 0, \\ -u & u < 0. \end{cases}$$

Hint. First show that the dual problem for general penalty function ϕ is

$$\begin{array}{ll} \text{maximize} & b^T\nu - \sum_{i=1}^m \phi^*(-\nu_i) \\ \text{subject to} & A^T\nu = 0 \end{array}$$

Then determine the specific conjugate function ϕ^* for (a) and (b) above.

4. **Estimation of mean and variance (30 marks).** Consider a random variable $x \in \mathbb{R}$ with density p, which is normalized, i.e., has zero mean and unit variance. Now consider a random variable $y = (x + b)/a$ obtained by an affine transformation of x, where $a > 0$. The random variable y has mean b/a and variance $1/a^2$. As a and b vary over $\mathbb{R}_+$ and $\mathbb{R}$, respectively, we generate a family of densities obtained from p by scaling and shifting, uniquely parametrized by mean and variance.

   (a) Show that if p is log-concave, then finding the maximum-likelihood estimates of a and b, given samples $y_1, \ldots, y_n$ of y, is a convex problem.

   (b) As an example, work out an analytical solution for the maximum-likelihood estimates of a and b, assuming p is a normalized Gaussian density, $p(x) = e^{-\frac{1}{2}x^2}$. You may consider first minimizing over b and then over a.