CS 680

Homework 1 (due September 18th)

1. **Ridge regression [10 pts].**
   The ridge regression optimization can be solved by gradient descent. Write this algorithm in the primal and dual formulations.

2. **A kernel for discrete data [10 pts].**
   Let Ω be a set, and let A, B be subsets of Ω. Show that the following functions are kernels:
   $$K_1(A, B) = |A \cap B|,$$
   and
   $$K_2(A, B) = 2^{|A \cap B|},$$
   where $|\cdot|$ is the cardinality of a set.

3. **Is this a kernel? [35 pts]**
   Let K_1 and K_2 be kernels over $\mathcal{X} \times \mathcal{X}$. Let $f : \mathcal{X} \mapsto \mathbb{R}$, let a be a positive number, let ϕ be a function $\phi : \mathcal{X} \mapsto \mathbb{R}^m$, and let K_3 be a kernel over $\mathbb{R}^m \times \mathbb{R}^m$. For each of the functions K defined below, state whether it is a kernel. If you think it is a kernel, prove it by explicitly showing the feature map for which K is a dot product; otherwise prove it is not a kernel.

   (a) $K(\mathbf{x}, \mathbf{x}') = K_1(\mathbf{x}, \mathbf{x}') + K_2(\mathbf{x}, \mathbf{x}')$
   (b) $K(\mathbf{x}, \mathbf{x}') = K_1(\mathbf{x}, \mathbf{x}') - K_2(\mathbf{x}, \mathbf{x}')$
   (c) $K(\mathbf{x}, \mathbf{x}') = aK_1(\mathbf{x}, \mathbf{x}')$
   (d) $K(\mathbf{x}, \mathbf{x}') = -aK_1(\mathbf{x}, \mathbf{x}')$
   (e) $K(\mathbf{x}, \mathbf{x}') = K_1(\mathbf{x}, \mathbf{x}')K_2(\mathbf{x}, \mathbf{x}')$
   (f) $K(\mathbf{x}, \mathbf{x}') = K_3(\phi(\mathbf{x}), \phi(\mathbf{x}'))$
   (g) $K(\mathbf{x}, \mathbf{x}') = f(\mathbf{x})f(\mathbf{x}')$
   (h) $K(\mathbf{x}, \mathbf{x}') = \frac{K_1(\mathbf{x}, \mathbf{x}')}{\sqrt{K_1(\mathbf{x}, \mathbf{x})K_1(\mathbf{x}', \mathbf{x}')}}$

4. **Positivity on the diagonal [5 pts].** Prove that every kernel satisfies $K(\mathbf{x}, \mathbf{x}) \geq 0 \; \forall \mathbf{x} \in \mathcal{X}$.

5. **Kernels with a vanishing diagonal [5 pts].** Prove that a kernel satisfying $K(\mathbf{x}, \mathbf{x}) = 0 \; \forall \mathbf{x} \in \mathcal{X}$ is identically zero. You can use the Cauchy-Schwarz inequality which states that for all $\mathbf{x}, \mathbf{x}'$ in some vector space $\mathcal{X}$
   $$|\langle \mathbf{x}, \mathbf{x}' \rangle| \leq ||\mathbf{x}|| \cdot ||\mathbf{x}'||,$$
   with equality occurring if and only if $\mathbf{x}$ and $\mathbf{x}'$ are co-linear (i.e. there exists λ such that $\mathbf{x}' = \lambda \mathbf{x}$).

6. **Positive definiteness does not imply positivity [5 pts].** Give an example of a kernel which is positive definite, but not positive, i.e. it does not satisfy $K(\mathbf{x}, \mathbf{x}') \geq 0 \; \forall \mathbf{x} \in \mathcal{X}$.

7. **Classifier performance as a function of kernel and classifier parameters [30 pts].** Download the dataset provided on the homework 1 section of the homework page of the course website. In this assignment we will explore the dependence of classifier accuracy on the kernel, kernel parameters, and classifier parameters, using the ridge regression classifier implemented in `PyML`. You can instantiate a ridge regression classifier as:

```
>>> from PyML import classifiers
>>> rr = classifiers.RidgeRegression(ridge = someValue)
```

Accuracy can be assessed using cross-validation:

```
>>> results = rr.cv(data)
```

where `data` is a dataset object (see the tutorial on how to read data into `PyML`). By default a dataset is instantiated with a linear kernel attached to it. To use a different kernel you need to attach a new kernel to the dataset:

```
>>> from PyML import ker
>>> data.attachKernel(ker.Gaussian(gamma = 2.0))
```

or

```
>>> from PyML import ker
>>> data.attachKernel(ker.Polynomial(degree = 3))
```

In this question we will consider both the Gaussian and polynomial kernels:

$$K_{gaus} = \exp(-\gamma ||\mathbf{x} - \mathbf{x}'||^2)$$

$$K_{poly} = (1 + \langle \mathbf{x}, \mathbf{x}' \rangle)^p$$

Plot the accuracy of the classifier, measured using the success rate and the area under the ROC curve as a function of both the ridge parameter of the classifier, and the free parameter of the kernel function. Show a couple of representative cross sections of this plot for a given value of the ridge parameter, and for a given value of the kernel parameter. Comment on the results.