Homework 6

Due April 5 at 11 pm

1. (Gradient descent and ridge regression) In this problem we study the iterations of gradient descent applied to the ridge-regression cost function

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2} \left| \left| y - X^T \beta \right| \right|_2^2 + \frac{\lambda}{2} \left| \left| \beta \right| \right|_2^2, \tag{1}$$

where $X \in \mathbb{R}^{p \times n}$ is a fixed feature matrix and $y \in \mathbb{R}^n$ is a response vector. (The factor of 1/2 is just there to make calculations a bit cleaner.)

- (a) Derive a closed form expression for the value of the estimated coefficient $\beta^{(k)}$ at the kth iteration of gradient descent initialized at the origin in terms of the SVD of X when the step size is constant.
- (b) Under what condition on the step size does gradient descent converge to the ridge-regression coefficient estimate as $k \to \infty$?
- (c) Assume the following additive model for the data:

$$\tilde{y}_{\text{train}} := X^T \beta_{\text{true}} + \tilde{z}_{\text{train}},$$
 (2)

where \tilde{z}_{train} is modeled as an *n*-dimensional iid Gaussian vector with zero mean and variance σ^2 . What is the distribution of the estimated coefficient $\tilde{\beta}^{(k)}$ at the *k*th iteration of gradient descent initialized at the origin?

- (d) Complete the script $RR_GD_landscape.py$ in order to verify your answer to the previous question. Report the figures generated by the script.
- 2. (Climate modeling) In this problem we model temperature trends using a linear regression model. The file t_data.csv contains the maximum temperature measured each month in Oxford from 1853-2014. We will use the first 150 years of data (the first 150 · 12 data points) as a training set, and the remaining 12 years as a test set.

In order to fit the evolution of the temperature over the years, we fit the following model

$$y[t] = a + bt + c\cos(2\pi t/T) + d\sin(2\pi t/T)$$
 (3)

where $a, b, c, d \in \mathbb{R}$, y[t] denotes the maximum temperature in Celsius during month t of the dataset (with t starting from 0 and ending at $162 \cdot 12 - 1$).

- (a) What is the number of parameters in your model and how many data points do you have to fit the model? Are you worried about overfitting?
- (b) Fit the model using least squares on the training set to find the coefficients for values of T equal to $1,2,\ldots,20$. Which of these models provides a better fit? Explain why this is the case. In the remaining question we will fix T to the value T^* that provides a better fit.

- (c) Produce two plots comparing the actual maximum temperatures with the ones predicted by your model for $T := T^*$; one for the training set and one for the test set.
- (d) Fit the modified model

$$y[t] = a + bt + d\sin(2\pi t/T^*) \tag{4}$$

and plot the fit to the training data as in the previous question. Explain why it is better to also include a cosine term in the model.

- (e) Provide an intuitive interpretation of the coefficients a, b, c and d, and the corresponding features. According to your model, are temperatures rising in Oxford? By how much?
- 3. (Sines and cosines) Let $x: [-1/2, 1/2) \to \mathbb{R}$ be a real-valued square-integrable function defined on the interval [-1/2, 1/2), i.e. $x \in L_2[-1/2, 1/2)$. The Fourier series coefficients of x, are given by

$$\hat{x}[k] := \langle x, \phi_k \rangle = \int_{-1/2}^{1/2} x(t) \exp\left(-i2\pi kt\right) dt, \quad k \in \mathbb{Z},$$
 (5)

and the corresponding Fourier series of order k_c equals

$$\mathcal{F}_{k_c}\{x\}(t) = \sum_{k=-k_c}^{k_c} \hat{x}[k] \exp(i2\pi kt).$$
 (6)

As we will discuss in class, this is a representation of x in a basis of complex exponentials. In this problem we show that for real signals the Fourier series is equivalent to a representation in terms of cosine and sine functions.

- (a) Prove that $\hat{x}[k] = \overline{\hat{x}[-k]}$ for all $k \in \mathbb{Z}$. [Hint: What is $\overline{e^{it}}$?]
- (b) Show that the Fourier series of x of order k_c can be written as

$$\mathcal{F}_{k_c}\{x\}(t) = a_0 + \sum_{k=1}^{k_c} a_k \cos(2\pi kt) + b_k \sin(2\pi kt),$$

for some $a_0, \ldots, a_k, b_1, \ldots, b_k \in \mathbb{R}$. [Hint: Group terms in $\mathcal{F}_{k_c}\{x\}(t)$ corresponding to $\pm k$ and use previous part. What is the real part of zw for $z, w \in \mathbb{C}$?]

- (c) Give expressions for the coefficients a_k, b_k for $k \geq 1$ from the previous part as real integrals. Interpret them in terms of inner products.
- (d) Suppose $x(t) = \cos(2\pi(t+\phi))$ for some fixed $\phi \in \mathbb{R}$. What are the Fourier coefficients of x?
- (e) Suppose that f is also even (i.e., x(-t) = x(t)). Prove that the Fourier coefficients are all real (i.e., that $\hat{x}[k] \in \mathbb{R}$ for all $k \in \mathbb{Z}$).