A Theory

Problem 11 (Linear Programs and $\varepsilon$-insensitive Loss, 8 Points)
Assume we have the following loss function

$$c(x, y, f(x)) = |y - f(x)|_\varepsilon$$
where $|\xi|_\varepsilon := \begin{cases} 
0 & \text{if } |\xi| \leq \varepsilon \\
\xi - \varepsilon & \text{if } \xi > \varepsilon \\
-\xi - \varepsilon & \text{otherwise}
\end{cases}$

1. Rewrite $|\xi|_\varepsilon$ as a linear optimization problem (analogous to the rewrite of $|\xi|$ which was discussed in the lecture). **Hint:** all you need to do is modify the constraints.

2. Rewrite the regularized risk functional for a linear model $f(x) = \langle w, x \rangle$ as a quadratic optimization problem with constraints. **Hint:** you only need to take care of the empirical risk term.

Recall that the regularized risk is given by

$$R_{\text{reg}}[f] = \frac{1}{m} \sum_{i=1}^{m} |y_i - f(x_i)|_\varepsilon + \frac{\lambda}{2} \|w\|^2$$

Problem 12 (Prior Probabilities, 12 Points)
Assume a prior probability $p(f) = c \exp(-\frac{1}{2}(\|f\|^2 + \|f'\|^2))$ on $[0, 2\pi]$ for suitably chosen $c$.

1. For the class of functions $\mathcal{F}$ given by

$$\mathcal{F} := \{ f | f(x) = \alpha_0 + \alpha_1 \cos x + \beta_1 \sin x \}$$

with $\alpha_i, \beta_i \in \mathbb{R}$ compute the normalization constant such that $\int_{\mathcal{F}} p(f) df = 1$.

2. Now assume the class

$$\mathcal{F} := \left\{ f | f(x) = \alpha_0 + \sum_{i=1}^{n} \alpha_i \cos(ix) + \beta_i \sin(ix) \right\} .$$

What is the value of the normalization constant in this case. Rewrite $p(f)$ directly in terms of the coefficients $\alpha_i$ and $\beta_i$.

3. Consider the series $f_n := \sin x + \frac{1}{n} \sin nx$. Show that the series $f_n$ converges to $\sin x$ for $n \to \infty$, yet that $p(f_n)$ does not converge to $p(\sin x)$.

4. **Bonus question:** Interpret the previous result.
**Problem 13 (Generalized Linear Models, 20 Points)** Let us assume a generalized linear model on $\mathbb{R}$ where $f$ is given by

$$f(x) = a + bx + \sum_{i=1}^{n} \alpha_i \exp(-(i-x)^2)$$

for $a, b, \alpha_i \in \mathbb{R}$ and we have squared loss, i.e.

$$c(x, y, f(x)) = \frac{1}{2}(y - f(x))^2.$$ 

1. Implement in C/MATLAB the algorithm that takes $(x_1, \ldots, x_m), (y_1, \ldots, y_m),$ and $n$ as an input and produces $a, b$ and $\alpha_i$ which minimize the empirical risk as an output. **Note:** take care of cases where $m < n$. You can use `pinv` in MATLAB.

2. Test your program on data generated by $y = f(x) + \xi$ where $f(x) = 1 + 2x + 3 \exp(-(3-x)^2) + 2 \exp(-(5-x)^2)$

  More specifically, draw $x$ from $[0, 10]$ and let $\xi$ be normally distributed with zero mean and variance $\sigma$. Plot the estimate of $f(x)$ on $[0, 10]$ for

  - $(n = 10, m = 5, \sigma = 0)$, $(n = 10, m = 20, \sigma = 0)$, $(n = 10, m = 50, \sigma = 0)$,
  - $(n = 10, m = 5, \sigma = 0.5)$, $(n = 10, m = 20, \sigma = 0.5)$, $(n = 10, m = 50, \sigma = 0.5)$.

  **Hint:** You can script the testing.

3. Now we introduce a regularization term via

$$\Omega[f] = \frac{1}{2} \left( a^2 + b^2 + \sum_{i=1}^{n} \alpha_i^2 \right)$$

to minimize

$$R_{\text{reg}}[f] = R_{\text{emp}}[f] + \lambda \Omega[f]$$

Modify your C/MATLAB from above such that the algorithm takes $(x_1, \ldots, x_m), (y_1, \ldots, y_m)$, $\lambda$, and $n$ as an input and produces $a, b$ and $\alpha_i$ which minimize the empirical risk as an output.

4. Test your program in the setting (2) for

- $(n = 10, m = 30, \sigma = 0.1, \lambda = 0.01)$, $(n = 10, m = 30, \sigma = 0.1, \lambda = 0.1)$,
- $(n = 10, m = 30, \sigma = 0.1, \lambda = 1)$, $(n = 10, m = 30, \sigma = 0.5, \lambda = 0.01)$,
- $(n = 10, m = 30, \sigma = 0.5, \lambda = 0.1)$, $(n = 10, m = 30, \sigma = 0.5, \lambda = 1)$.

($n$: number of exponential terms, $m$: number of observations, $\sigma$: variance of additive noise, $\lambda$: regularization constant)