1 Probability Theory

Problem 1 [2 points] Your friend Mark Z. has a hepatic carcinoma checkup using a new medical technology. The result shows that Mark has the disease. The test is 98% accurate. In other words, this method detects hepatic carcinoma in 98 out of 100 times when it is there, and in 2/100 misses it; this method detects hepatic carcinoma in 2 of 100 cases when it is not there, and 98 of 100 times correctly returns a negative result. The previous research of hepatic carcinoma suggests that one in 2,000 people has this disease. What is the probability that Mark actually has hepatic carcinoma? Put down everything up to the point where you would use your calculator and stop there.

2 Linear Regression

Problem 2 [2 points] We have 1D input points $\mathbf{X} = [0, 1, 2]$, and corresponding 2D output $\mathbf{Y} = [\{-1, 1\}, \{1, -1\}, \{2, -1\}]$. We embed x_i into 2d with the basis function:

$$\Phi(0) = (1, 0)^T, \Phi(1) = (1, 1)^T, \Phi(2) = (2, 2)^T$$

The model becomes $\hat{\mathbf{y}} = \mathbf{W}^T \Phi(x)$. Compute the MLE for $\mathbf{W}$.

Put down everything up to the point where you would use your calculator and stop there.

Problem 3 [3 points] Assume that $\bar{\mathbf{x}} = 0$. We have ridge regression, and we minimize:

$$J(\mathbf{w}) = (\mathbf{y} - \mathbf{X}\mathbf{w} - w_0\mathbf{1})^T(\mathbf{y} - \mathbf{X}\mathbf{w} - w_0\mathbf{1}) + \lambda \mathbf{w}^T\mathbf{w}$$

Derive the optimizer

$$\hat{\mathbf{w}}_{ridge} = (\lambda \mathbf{I} + \mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

3 Logistic Regression

Consider the data in the following figure, where we fit the model $p(y = 1|\mathbf{x}, \mathbf{w}) = \sigma(w_0 + w_1x_1 + w_2x_2)$. Suppose we fit the model by maximum likelihood, i.e., we minimize

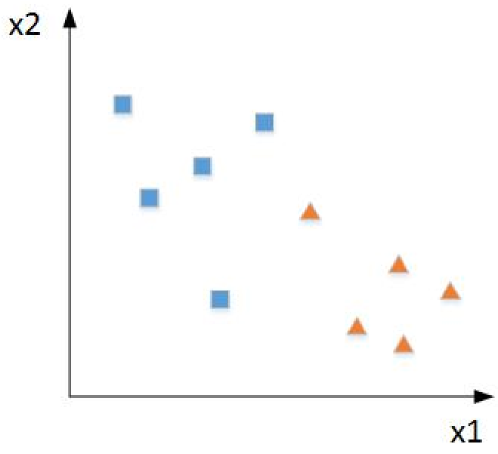
$$J(\mathbf{w}) = -L(\mathbf{w}, \mathcal{D}_{train})$$

where $L(\mathbf{w}, \mathcal{D}_{train})$ is the log likelihood on the training set.

Problem 4 [2 points] Suppose we regularize only the w_0 parameter, i.e., we minimize

$$J_0(\mathbf{w}) = -L(\mathbf{w}, \mathcal{D}_{train}) + \lambda w_0^2$$

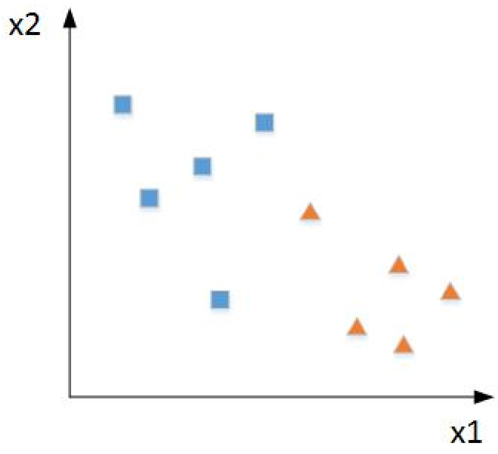
and λ is a very large number. Sketch a possible decision boundary. Show your work.



Problem 5 [2 points] Now suppose we heavily regularize only the w_1 parameter, i.e., we minimize

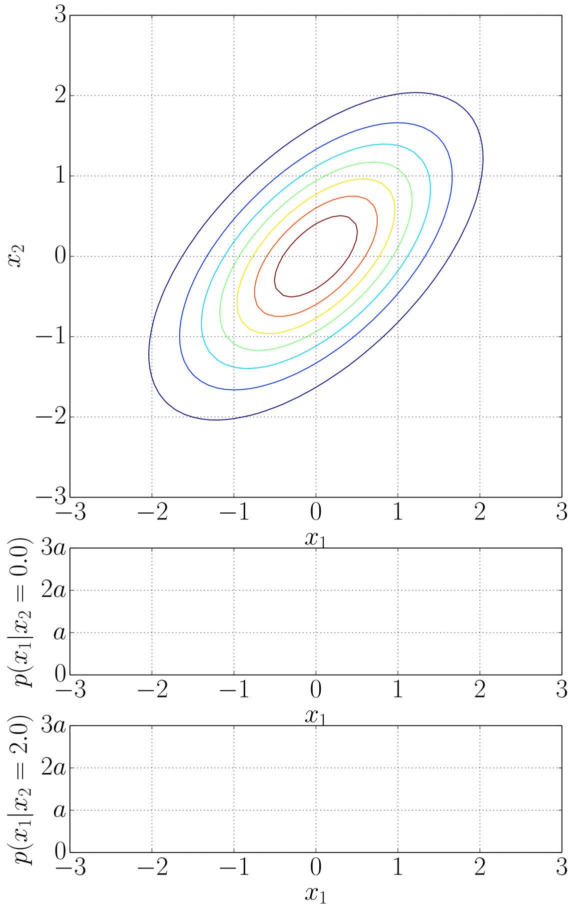
$$J_1(\mathbf{w}) = -L(\mathbf{w}, \mathcal{D}_{train}) + \lambda w_1^2$$

Sketch a possible decision boundary. Show your work.



4 Multivariate Gaussian

Problem 6 [2 points] The plot below shows a joint Gaussian distribution $p(x_1, x_2)$. Qualitatively (!) draw the conditionals $p(x_1|x_2 = 0)$ and $p(x_1|x_2 = 2)$ in the given coordinate systems. Note that the scaling a is arbitrary but fixed.



5 Kernels

We have $\mathbf{x} = [x_1\ x_2]^T$. Given the mapping

$$\varphi(x) = [1 \quad x_1^2 \quad \sqrt{2}x_1x_2 \quad x_2^2 \quad \sqrt{2}x_1 \quad \sqrt{2}x_2]^T$$

Problem 7 [2 points] Determine the kernel $K(\mathbf{x}, \mathbf{y})$. Simplify your answer.

6 Constrained optimization

Find the box with the maximum volume which has surface area no more than $S \in \mathbb{R}^+$.

Problem 8 [2 points] Derive the Lagrangian of the problem and the corresponding Lagrange dual function. Hint: set the parameters of the length, width and height to be l, w, h respectively.

Problem 9 [3 points] Solve the dual problem and give the solution to the original problem.

7 Neural Networks

This is an unfair mock question about material you have not yet seen.

Problem 10 [2 points] What is deep about learning? Describe in 2–3 sentences.

Problem 11 [2 points] What vanishes in the gradient?