HW Assignment 1, Theory (Due by 5:30pm on Feb 17)

1 Theory (60 points)

1. [**Derivatives**, 20 points]
   Compute the derivative $f'(x) = \dfrac{\delta f}{\delta x}$ or the gradient vector $\left[\dfrac{\delta f}{\delta x}, \dfrac{\delta f}{\delta y}\right]$ for the following functions $f(x)$ or $f(x, y)$. Make sure you show all your work (derivation steps):

$$
\begin{aligned}
f(x) &= 3x + 4 && (1)\\
f(x) &= 3x^2 + 4x + 2 && (2)\\
f(x) &= 3x^n + 4x && (3)\\
f(x) &= e^{2x} + 3x && (4)\\
f(x) &= \ln(x + 1) && (5)\\
f(x) &= \ln(2x^3 + 1) && (6)\\
f(x) &= 1/(x^2 + 1) && (7)\\
f(x, y) &= x^2 + 2xy + y^2 && (8)\\
f(x, y) &= 2(3x + 2y + 1)^2 && (9)\\
f(x, y) &= \ln(1 + e^{2x+3y}) && (10)
\end{aligned}
$$

2. [**Simple linear regression**, 20 points]
   Consider the problem of fitting a dataset of N points with using simple linear regression, by minimizing the sum-of-squares error:

$$J(\mathbf{w}) = \frac{1}{2N} \sum_{n=1}^{N} (h_{\mathbf{w}}(x_n) - t_n)^2 \qquad (11)$$

   where $h_{\mathbf{w}}(x_n) = w_1 x_n + w_0$. By setting the gradient of $J(\mathbf{w})$ to 0, prove that the solution to minimizing $J(\mathbf{w})$ satisfies the following set of linear equations:

$$
\begin{aligned}
w_0 N + w_1 \sum_{n=1}^{N} x_n &= \sum_{n=1}^{N} t_n && (12)\\
w_0 \sum_{n=1}^{N} x_n + w_1 \sum_{n=1}^{N} x_n^2 &= \sum_{n=1}^{N} t_n x_n && (13)
\end{aligned}
$$

3. [**Probability Theory**, 20 points] Suppose that we have three coloured boxes r (red), b (blue), and g (green). Box r contains 3 apples, 4 oranges, and 3 limes, box b contains 1 apple, 1 orange, and 0 limes, and box g contains 3 apples, 3 oranges, and 4 limes. If a box is chosen at random with probabilities $p(r) = 0.2$, $p(b) = 0.2$, $p(g) = 0.6$, and a piece of fruit is removed from the box (with equal probability of selecting any of the items in the box), then what is the probability of selecting an apple? If we observe that the selected fruit is in fact an orange, what is the probability that it came from the green box?

4. [**Ridge Regression (*)**, 20 bonus points]
Consider the regularized linear regression objective shown below:

$$J(\mathbf{w}) = \frac{1}{2N} \sum_{n=1}^{N} (h(\mathbf{x}_n, \mathbf{w}) - t_n)^2 + \frac{\lambda}{2} ||\mathbf{w}||^2$$

(a) Minimizing the L2 norm of $\mathbf{w}$ drives all parameters, including w_0, towards 0. Are there situations in which we do not want to constrain w_0 to be small? If yes, give an example, if not show why it is useful to constrain all the weights to be small, including w_0 .

(b) We have seen in class how to compute the weights $\mathbf{w}$ that minimize $J(\mathbf{w})$. Assume now that we replace $||\mathbf{w}||$ in $J(\mathbf{w})$ with $||\mathbf{w}_{[\mathbf{1}:]}||$, where $\mathbf{w}_{[\mathbf{1}:]} = [w_1, w_2, ..., w_M]$. Derive the solution for $\mathbf{w}$ that minimizes this new objective function.

2 Submission

Submit your responses on Canvas as one file named `theory.pdf`. You can use an editor such as Latex or Word that allows editing of equations. Alternatively, if you choose to write your solutions on paper, submit an electronic scan / photo of it into one file called `theory.pdf`. Make sure that your writing is legible and easy to read.

Clear and complete explanations and proofs of your results are as important as getting the right answer.