Collaborating on homework is encouraged, but you must write your own solutions in your own words and list your collaborators. Copying someone else's solution will be considered plagiarism and may result in failing the whole course.

Please answer clearly and concisely. Explain your answers. Unexplained answers will get lower scores or even no credits.

(1) (30 points) Expressiveness of concept classes
  (a) Prove that every 1-decision list can be represented as a linear threshold function.
  (b) Prove that every 1-decision list can be represented as an n-term DNF formula.

(2) (15 points) Consider the following variant of Winnow2 that doubles the weights even when not making mistakes on positive samples:

Winnow'

Initialize: $w_1 = \cdots = w_n = 1$, θ fixed to be n
On input x, output hypothesis $h(x) = \mathbb{1}(w \cdot x \geqslant \theta)$ and get $c(x)$
False positive ($h(x) = 1, c(x) = 0$): For every i s.t. $x_i = 1$
  Halve w_i (demotion)
Positive ($c(x) = 1$)**:** For every i s.t. $x_i = 1$
  Double w_i (promotion)

Prove that when learning monotone disjunctions in the online model, this algorithm may make an arbitrarily large number of mistakes.

(3) (15 points) Let $\mathcal{C}_W$ denote the collection of linear threshold functions $c(x) = \mathbb{1}(w \cdot x \geqslant \theta)$ over instance space $X = \{0,1\}^n$ such that the weights w_i are all nonnegative integers and $\sum_{1 \leqslant i \leqslant n} w_i \leqslant W$.
Consider learning $\mathcal{C}_W$ in the Online model using Winnow2 or Perceptron algorithms.
  (a) How many mistakes does Winnow2 make?
  (b) Show that Perceptron makes $O(W^2 n)$ mistakes.
Justify how you arrive at the above mistake bounds.

(4) (30 points) VC dimension of concept classes
  (a) Let $X = \{0,1\}^n$ and $\mathcal{C}$ be the collection of 1-sparse monotone disjunctions on X (i.e. monotone disjunctions that involve exactly 1 variable).
  Prove that $\mathrm{VCDim}(\mathcal{C}) = \lfloor \log n \rfloor$.
  (b) A general parity function is of the form $c(x) = x_{i_1} \oplus x_{i_2} \oplus \cdots \oplus x_{i_k} \oplus b$.
  Here $b \in \{0,1\}, 1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n$ and $0 \leqslant k \leqslant n$. The symbol $\oplus$ denotes addition mod 2.
  Let $X = \{0,1\}^n$ and $\mathcal{C}$ be the collection of general parity functions on X. Prove that $\mathrm{VCDim}(\mathcal{C}) = n + 1$.

(5) (10 points) Almost every algorithm in the online mistake bound model in this course (Winnow, Perceptron, decision list, etc) updates its hypothesis only after making a mistake. If the algorithm doesn't make a mistake, its hypothesis is not updated.
We want to show this behaviour holds without loss of generality:
Suppose some deterministic algorithm A learns a concept class $\mathcal{C}$ in the Online Learning model making at most M mistakes, and A may change its hypothesis even when not making a mistake. Show how to modify A to get a deterministic algorithm B that only updates its hypothesis after a mistake, and B still makes at most M mistakes.