A Classification Algorithm is a procedure for selecting a hypothesis from a set of alternatives that best fits a set of observations.
Distance from person to the food

Size of the person

+ = Got the food
- = Got caught
Distance from person to the food

Size of the person

+ = Got the food
- = Got caught
Training Set -- Set of labeled examples
   (Set of observations begin matched by the learning algorithm.)

Labeled Example -- Vector of feature values and associated label.
   (A single observation)

Feature Value -- A numeric or ordinal value.
   (The value of one particular aspect of an observation)

Label -- A numeric or ordinal value.
   (The value to be predicted by the learning algorithm.)

Hypothesis -- A mapping from feature vectors to labels
   (The output of the learning algorithm. One possible pattern observed in the data.)

Hypothesis Space -- A set of hypotheses.
   (The space of all possible hypotheses considered by a learning algorithm.)

Target Concept -- A mapping from feature vectors to labels.
   (The "real world" mapping being sought. We assume the training set is generated from a distribution base upon this mapping. We do not assume the target concept is a member of the hypothesis space.)

Objective Function -- A mapping that maps a training set and hypothesis to a real value.
   (Measures the match between a set of observations and a hypothesis that seeks to predict those observations. Also called the loss function, cost function, or utility measure.)

Classification Algorithm -- A mapping from training sets to hypotheses that minimizes the objective function.
   (Supervised learning algorithm. Most practical classification algorithms are approximate, they strive to but go not guarantee to return the hypothesis that minimized the objective function.)

Components of a Classification Learning System
   (input/output of Supervised Learning)
Hypothesis Space
Choice

Training Set → Classification Algorithm → Induced Hypothesis

Testing Set → Testing → Score

Block Diagram of a Classification Algorithm
Distance across front of house

Distance from house

+ = Relief in safety
- = Scolded by owner

Relief in the front yard
Distance across front of house

Distance from house

+ = Relief in safety
- = Scolded by owner

Relief in the front yard
Distance across front of house

Distance from house

+  = Relief in safety
-
  = Scolded by owner

Relief in the front yard
Distance across front of house
Distance from house

+ = Relief in safety
- = Scolded by owner

Relief in the front yard
Distance across front of house

Distance from house

Relief in safety

Scolded by owner

Relief in the front yard
Distance across front of house

Distance from house

+ = Relief in safety
-
  = Scolded by owner

Relief in the front yard
Generalization -- Making predictions about unseen examples based on observations

More Hypotheses = Less Generalization
Distance from person to the food

Size of the person

+ = Got the food
-
= Got caught
ROC -- Receiver Operator Characteristic
Bias -- Everything that affects the hypothesis selected by the Classification Algorithm except the training data itself.

Bias -- A measure (in terms of the objective function) of the inherent error of the classification algorithm, given unbounded amounts of training data. (It is a measure of the "best the algorithm can do")
Two Hypothesis Spaces S1 and S2
The variance between hypotheses
Distance across front of house

Distance from house

Variance in Relief
Weak Bias
Large Hypothesis Space
Large Training Set
Overfit
Variance Error
Low Stability
Noise Sensitive

Strong Bias
Small Hypothesis Space
Small Training Set
Overgeneralization
Bias Error
High Stability
Noise Tolerant
The VC-Dim of a set of hypotheses is the largest N such that there exists a set of N feature vectors shattered by the hypothesis space.

A set of feature vectors is *shattered* by a set of hypotheses iff there exists a hypothesis H for each possible binary labeling of those vectors.
+ Ex01  0 0 1 1 0 0 0 1 1 0 0 0 1 0 0 0 1 0 0
+ Ex02  0 0 1 1 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0
+ Ex03  0 0 0 1 1 0 0 0 0 1 1 0 0 1 0 1 0 0 0
+ Ex04  1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
- Ex05  0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 1 0 0
- Ex06  0 0 1 0 0 0 1 0 0 0 0 1 0 1 0 0 0 1 0
- Ex07  0 0 0 0 1 0 0 1 1 0 1 1 0 0 1 0 1 0 0
- Ex08  0 0 1 0 0 1 0 0 1 1 1 0 1 0 1 0 1 0 0

x=+ if  a+d>1  or  c+h-l>1  ...
Distance to nearest flower

+ = Relief in safety
-

= Scolded by owner

Re-Representing the front yard problem
Problem 1: Fit the problem of recognizing alphanumeric characters into the classification framework. Specify with two sentences or a paragraph what each of the framework's components are in the case of character recognition, and how they would be obtained.

Problem 2: Propose a variant with a stronger bias by specifying a different representation of the examples.

Problem 3: Again propose another variant that has a stronger bias, but in this case specify a different hypothesis space (leaving the representation fixed.)