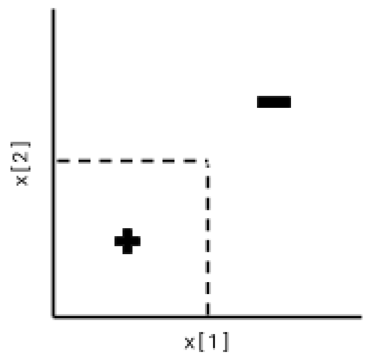
Solution 1



Solution 2

Minimum weighted error = 0.2 is for income.

$\hat{W}_t = \frac{1}{2}\ln(\frac{1-0.2}{0.2}) = 0.69$

Denoting old weight by α_i, the new weights are:

$\tilde{\alpha}_i = \alpha_i e^{\hat{W}_t} = \alpha_i e^{0.69} = \alpha_i/2$, if $\hat{y}_i = f_t(x_i) = y_i$ and

$\tilde{\alpha}_i = \alpha_i e^{-\hat{W}_t} = \alpha_i e^{-0.69} = 2\,\alpha_i$, if $\hat{y}_i = f_t(x_i) \neq y_i$.

New weights:

0.25	0.75	3	1	2	1.25	1.5	1	1	2	1

Solution 3

For all the models we've looked at so far (Linear Regression, Ridge/LASSO Regression, Logistic Regression, and Decision Trees), Write down the following:

1. Training Data:
    a. Linear, LASSO, Ridge: $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ for $\mathbf{x} \in \mathbb{R}^d, y \in \mathbb{R}$
    b. Logistic, Decision: $\{\mathbf{x}_i, y_i)\}_{i=1}^n$ for $\mathbf{x}^{(j)}$ either continuous or categorical, $y \in \{-1, +1\}$
2. Feature Extraction: What types of features can we extract for the different models, what methods have we talked about for feature extraction.
    a. Linear, LASSO, Ridge, Logistic: Any function of the data inputs. If we want to use categorical variables, we need to use one hot encoding.

   b. Decision Trees: Can also use any transformation of the data inputs. Does not require modifying categorical features.
3. ML Model: What are the parameters/weights we learn?
   a. Linear, LASSO, Ridge, Logistic: $\hat{\mathbf{w}}$ of weights where $\hat{f}(\mathbf{x}) = \hat{\mathbf{w}}^T\mathbf{x}$
   b. Decision Trees: An instance of a tree with splits.
4. ML Algorithm: What methods are used to compute the model?
   a. Linear, LASSO, Ridge: Gradient descent on the loss function
   b. Logistic: Gradient ascent
   c. Decision Tree: Try on all splits
5. Quality Metric: What metrics do we use to measure our models?
   a. Linear: RSS, MSE, RMSE
   b. LASSO: Same as linear, plus an L1 norm on the weights
   c. Ridge: Same as linear, plus an L2 norm on the weights
   d. Logistic: $$l(w) = \prod_{i=1}^{N} P(y_i|\mathbf{x}_i, w)$$
   e. Decision Tree: Classification error

Compare and contrast the two classification methods we've discussed so far: Decision Trees and Logistic Regression. Some questions to get you started:

1. What features does each method use? Both take advantage of categorical and continuous inputs. In order to cross correlate terms, or learn more complex decision boundaries, features combining inputs or exponentiating inputs needs to be explicitly done when choosing features. Decision trees can handle correlating inputs natively with infinite splits.
2. How do the quality metrics compare? In both instances we'd like to use classification error as the metric we optimize. But as we discussed in class, the classification error is not optimizable, so we use the likelihood. For decision trees, we can use the classification error as our quality metric, and we split on accuracy at each of the split nodes.
3. How do each handle continuous and categorical variables? For logistic regression, we need to use one hot encoding (dummy variables) to train our regression model on the data. Decision trees are able to split on categorical variables "out of the box".
4. How do each prevent overfitting? Logistic regression can use the same regularization techniques talked about with linear regression. For decision trees we talked about pruning. Also with bagging (Random forests) and boosting models (AdaBoost), we can overcome overfitting by using many instances of a decision tree model.