CS434 — sample questions

1. Short questions.
   - Explain why replacing the linear dot product with a non-linear kernel function (e.g., the quadratic kernel) in SVM allows it to learn a non-linear decision boundary.
   - Consider the following strategy for selecting $k$ for $k$-means. Run $k$-means using $k=2, 3, \cdots, k_{max}$, choose the $k$ value that gives the lowest Sum of Squared Error (SSE). What is wrong with this strategy?
   - In perceptron, let $w_t$ denote the current weight vector and it misclassifies $(x_t, y_t)$. What is the update rule for creating the new weight $w_{t+1}$? Explain why this update rule can correct for the mistake on $(x_t, y_t)$.
   - Please indicate for each action below whether it increases, decreases, or does not impact overfitting.
     * For KNN, change $k$ from 5 to 1
     * Change parameter $c$ for soft-margin SVM from 1000 to 0.1
     * Change from a linear Kernel to a 3-rd order polynomial kernel in SVM.

2. True or false questions.
   - (True or False) Logistic regression learns a non-linear decision boundary because it assumes that $p(y|X) = \frac{1}{1+e^{-w \cdot X}}$, which is a nonlinear function of $X$.
   - (True or False) When learning a linear decision boundary with the perceptron algorithm, it is guaranteed to converge within a finite number of steps.
   - (True or False) When applying $k$-means algorithm to cluster a given set of data, it is guaranteed to converge within a finite number of steps.
   - (True or False) If random variable A and B are independent from one another, they must be conditionally independent given C, where C can be any arbitrary random variable.
• We want to predict the stock price of an S&P 500 company. According to the recent KDNugget news, the perfect set of variables are: Butter production in Bangladesh (B), U.S. and Bangladesh Cheese production (C), and the Sheep population in U.S. and Bangladesh (S). We collected the following training data.

<table>
<thead>
<tr>
<th>B</th>
<th>C</th>
<th>S</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>low</td>
<td>low</td>
<td>up</td>
</tr>
<tr>
<td>high</td>
<td>high</td>
<td>low</td>
<td>up</td>
</tr>
<tr>
<td>high</td>
<td>low</td>
<td>high</td>
<td>down</td>
</tr>
<tr>
<td>low</td>
<td>low</td>
<td>low</td>
<td>down</td>
</tr>
<tr>
<td>low</td>
<td>high</td>
<td>high</td>
<td>up</td>
</tr>
</tbody>
</table>

3 Please use the Naive Bayes classifier to make prediction for (B=low, C=low, S=high). Please show your steps, which should clearly state all the distributions you use for your Naive Bayes classifier. A simple answer of predicting up or down will not be given points.

– Suppose now we learned that the price always goes down when $B = S$ and goes up when $B \neq S$. Does the Naive Bayes assumption hold and why?

– Build a depth-one decision tree (a.k.a a decision stump) using error-rate as the uncertainty measure. Please show your steps and clearly mark out the class labels for the leaf nodes of your decision tree.

– Can the top-down greedy algorithm learn a decision tree to correctly represent the class concept described in (b)? Why?