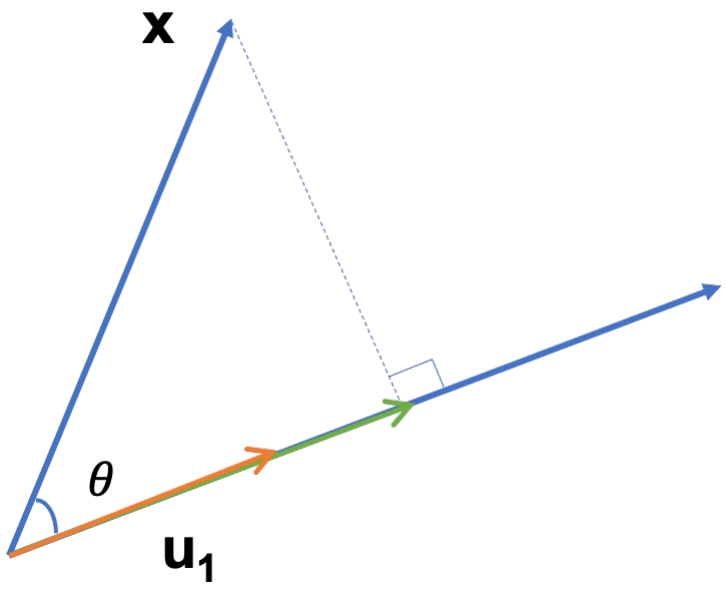
Homework 5 (Written)

Problem 1

As the figure shown below, there are two vectors $\boldsymbol{x}$ and $\boldsymbol{u}_1$ (orange color), where $\boldsymbol{u}_1$ has unit length. The projection of $\boldsymbol{x}$ onto $\boldsymbol{u}_1$ is represented by the vector in green.



Show that the length of the projection of $\boldsymbol{x}$ onto $\boldsymbol{u}_1$ has length $\boldsymbol{u}_1^T\boldsymbol{x}$.

Problem 2

Show that

$$\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{u}_1^{\mathrm{T}}\mathbf{x}_i - \mathbf{u}_1^{\mathrm{T}}\overline{\mathbf{x}}\right)^2 = \mathbf{u}_1^{\mathrm{T}}\mathbf{S}\mathbf{u}_1$$

where S is the data covariance matrix defined by

$$\mathbf{S} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{x}_i - \overline{\mathbf{x}}) (\mathbf{x}_i - \overline{\mathbf{x}})^{\mathrm{T}}$$

that is, Eqn (2) holds with S defined in Eqn. (3) in Lecture 9.

Problem 3

For a binary classification problem with three binary features, the data is shown below:

a	b	c	K
1	0	1	1
1	1	1	1
0	1	1	0
1	1	0	0
1	0	1	0
0	0	0	1
0	0	0	1
0	0	1	0

(a) Based on the naive Bayes classifier, what is $P(K = 1|a = 1, b = 1, c = 0)$?

(b) Based on the naive Bayes classifier, what is $P(K = 0|a = 1, b = 1)$?