

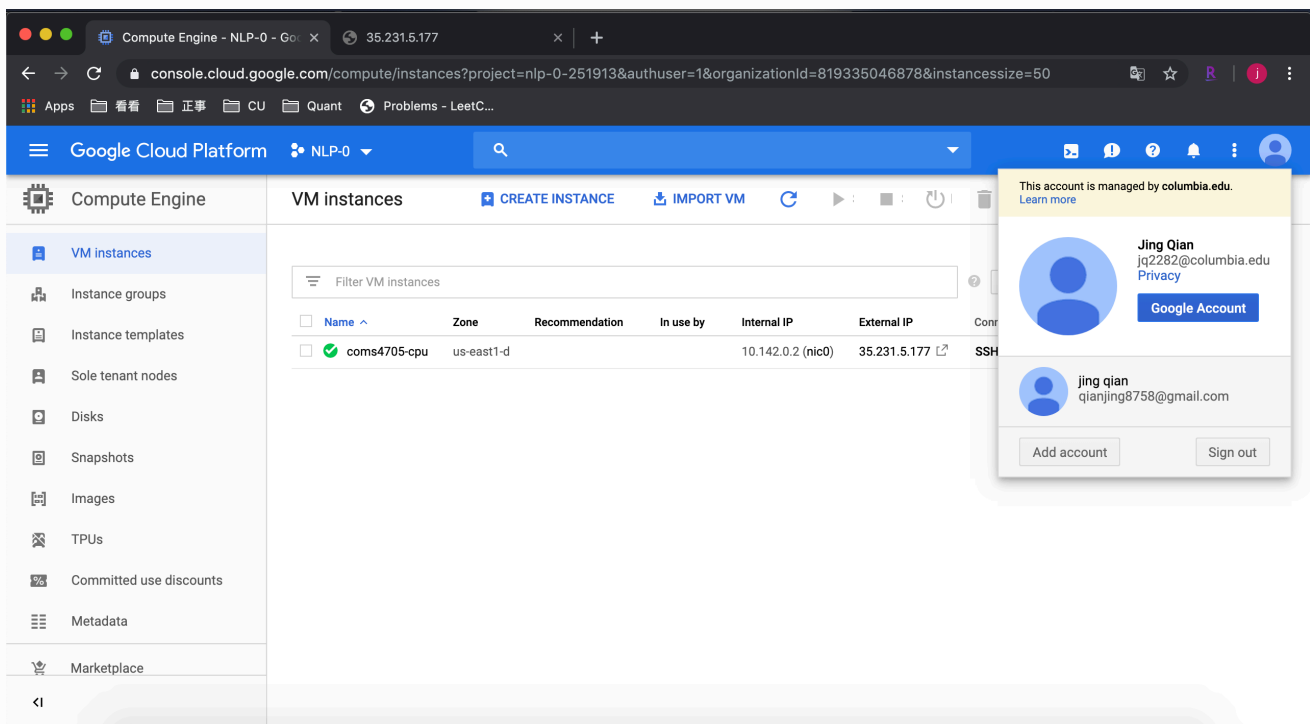
HOMEWORK 0 (COMS W4705)

Jing Qian (jq2282@columbia.edu)

1. Environment Setup and Programming

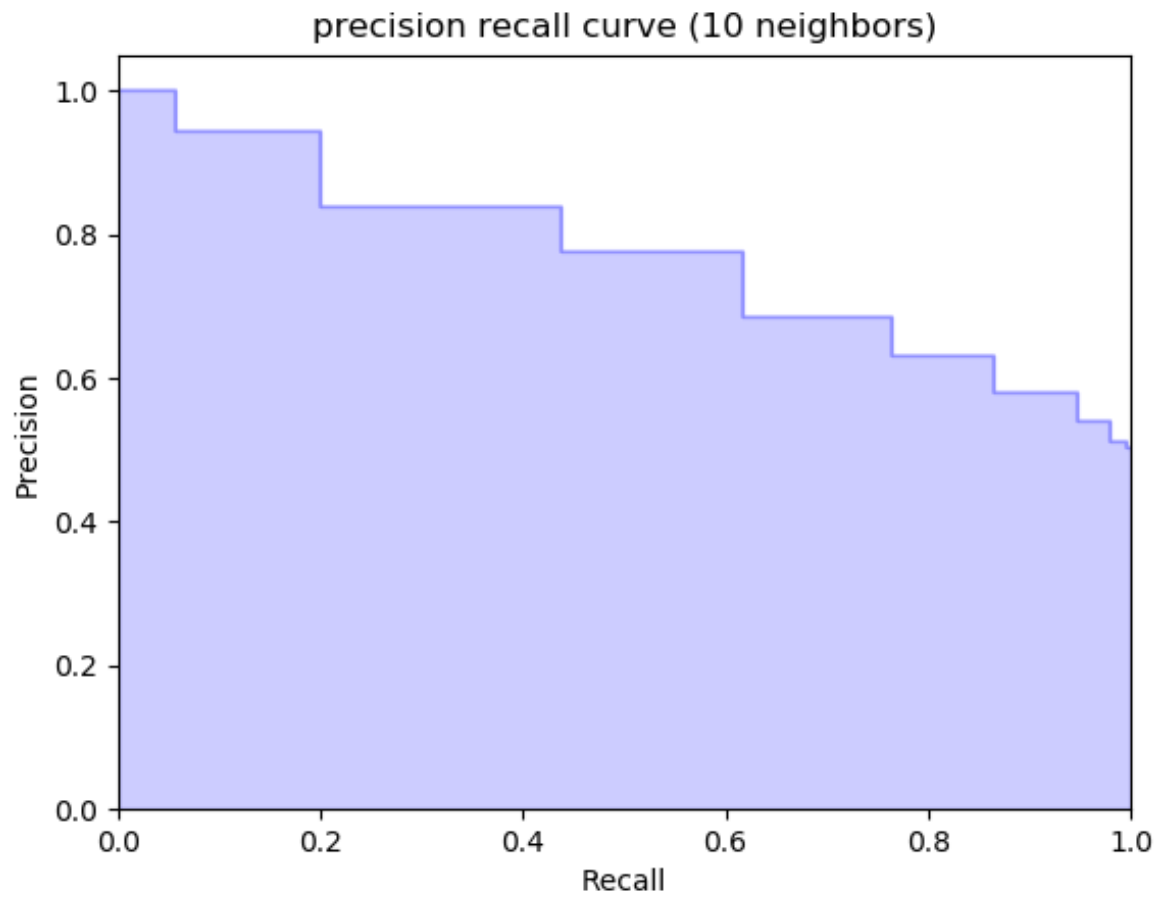
1.1 Enviroment Setup

Following is the screenshot of my Google Cloud "VM instances" page showing my virtual machine running.

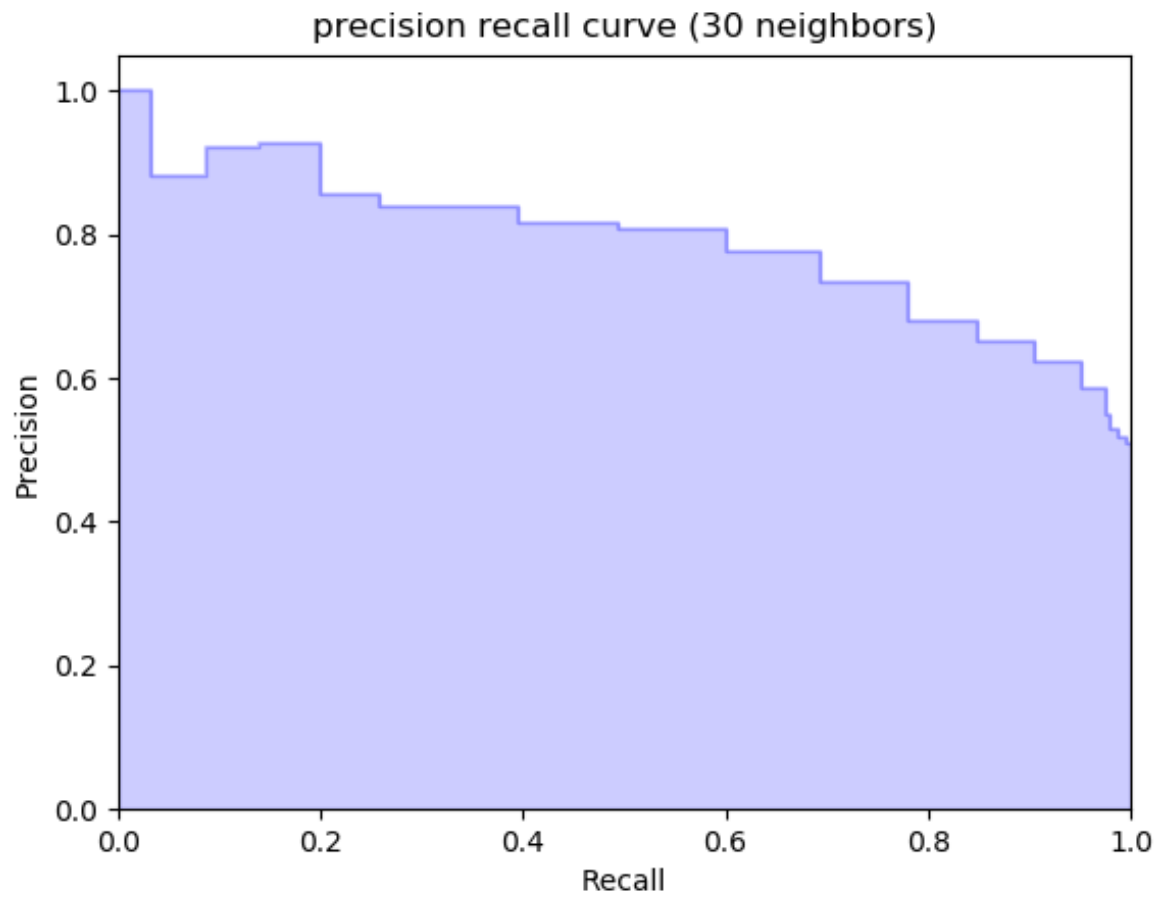


1.2 Programming

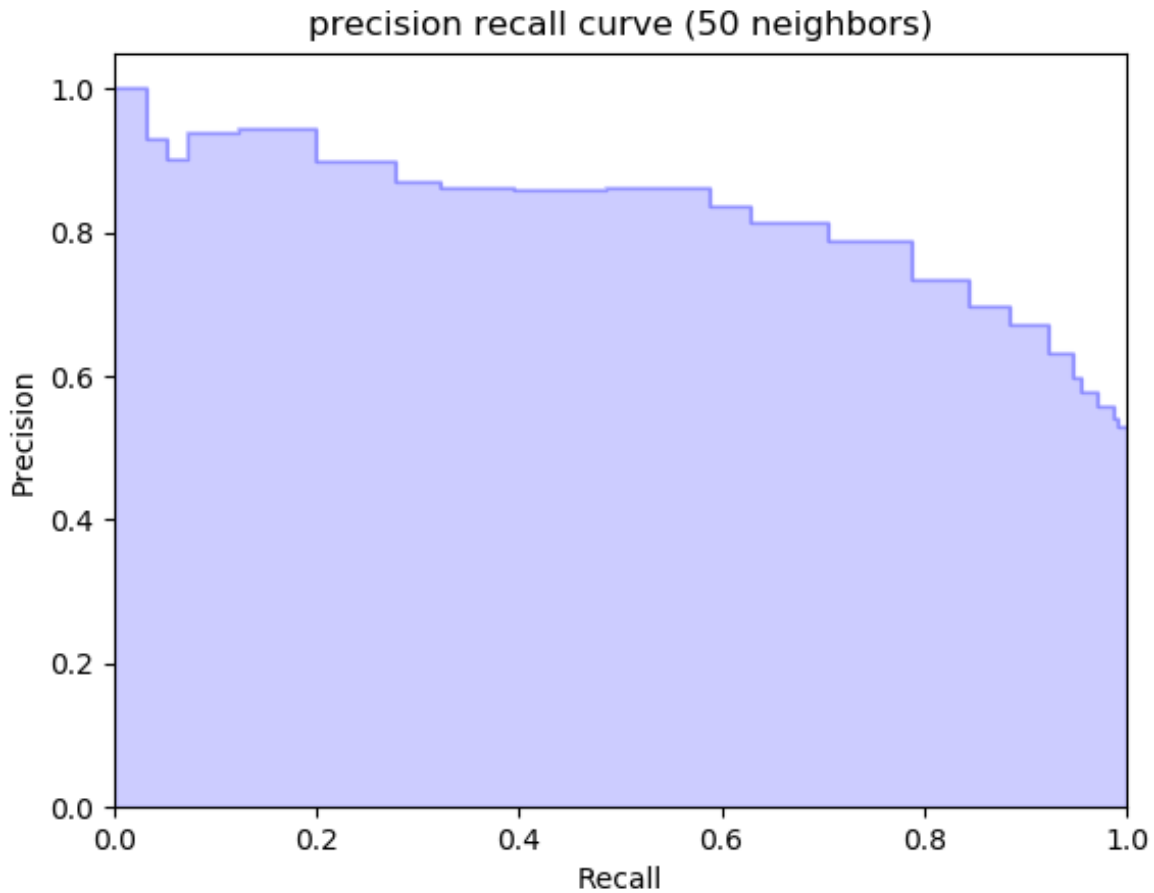
- Plot the precision-recall curve for the existing classifier:



- Plot the precision-recall curve when the number of neighbors is 30 instead.



- Plot the precision-recall curve when the number of neighbors is 50 instead.



需要写话？

2. Calculus

2.1 Chain rule and multivariate derivatives

i) From the functions provided, we have:

$$g(x, y) = x^2y - xh(x^2, y) = x^2y - x(x^2y^2 + 5) = x^2y - x^3y^2 - 5x$$

So:

$$\frac{\partial f}{\partial x} = g + x \frac{\partial g}{\partial x} = (x^2y - x^3y^2 - 5x) + x(2xy - 3x^2y^2 - 5) = 3x^2y - 4x^3y^2 - 10x$$

$$\frac{\partial f}{\partial y} = x \frac{\partial g}{\partial y} + 2 = x(x^2 - 2x^3y) + 2 = -2x^4y + x^3 + 2$$

ii)

$$\begin{aligned}\frac{\partial f}{\partial x} &= 1/y^2 + z \exp(x^2)(2x) = 1/y^2 + 2xz \exp(x^2) \\ \frac{\partial f}{\partial y} &= x(-2)/y^3 = -2x/y^3 \\ \frac{\partial f}{\partial z} &= \exp(x^2)\end{aligned}$$

2.2 Maxima and minima

From the expression of $f(x)$, we know that $f(x)$ is symmetrical about $x = 1/2$. At the upper bound 0 and lower bound 1 of the domain of x , $f(x)$ is zero, i.e., $f(0) = f(1) = 0$.

We get the first and second derivatives of $f(x)$ as following:

$$\begin{aligned}f'(x) &= \log_2 x + x \frac{1}{x \ln 2} - \log_2(1-x) + (1-x) \frac{1}{-(1-x) \ln 2} = \log_2 \frac{x}{1-x} \\ f''(x) &= \frac{1}{\frac{x}{1-x} \ln 2} \frac{(1-x) + x}{(1-x)^2} = \frac{1}{x(1-x) \ln 2}\end{aligned}$$

For $x \in (0, 1)$, the first and second derivatives of $f(x)$ are continuous, which could help us find the maxima and minima.

When $x = 1/2$, $f'(x) = 0$ and $f''(x) > 0$. So the **minima** of $f(x)$ for $x \in [0, 1]$ is $f(1/2) = -1$. Since $f(x)$ is symmetrical about $x = 1/2$, which is also the minima point and $f''(x)$ is positive over $x \in (0, 1)$, the **maxima** of $f(x)$ for $x \in [0, 1]$ is $f(0) = f(1) = 0$.

3. Probability and Statistics

3.1 Conditional probability

Here we use $P(b)$ to denote the probability of the first strip is "buffalo", and hence $P(b, b|b)$ denotes the probability that the second strip is also "buffalo" given the first strip is "buffalo". Then the probability that we pull out the following words "buffalo buffalo buffalo" is:

$$P(b, b, b) = P(b)P(b, b|b)P(b, b, b|b, b) = \frac{5}{10} * \frac{5-1}{10-1} * \frac{5-2}{10-2} = \frac{1}{12}$$

3.2 Bayes' rule

Here event X is the fact that I get a text from Maria about dogs, event Y is that the sender is Maria B and event Z is that the sender is Maria A. So using the Bayes' rule to calculate the probability this dog-content message is from Maria is:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} = \frac{P(X|Y)P(Y)}{P(X|Y)P(Y) + P(X|Z)P(Z)} = \frac{90\% * 50\%}{90\% * 50\% + 10\% * 50\%} = 90\%.$$

4 Linear Algebra

4.1 Basic matrix operations

i) According to the definition of the matrix multiplication, the matrix product $C_{n \times p} = A_{n \times m} B_{m \times p}$ has element calculated as $c_{ij} = \sum_{k=1}^m a_{ik} b_{kj}$. So we have:

$$\begin{bmatrix} 0 & 2 \\ 1 & 3 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 5 & 2 \\ 4 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 0*1 + 2*4 & 0*5 + 2*3 & 0*2 + 2*1 \\ 1*1 + 3*4 & 1*5 + 3*3 & 1*2 + 3*1 \\ 2*1 + 0*4 & 2*5 + 0*3 & 2*2 + 0*1 \end{bmatrix} = \begin{bmatrix} 8 & 6 & 2 \\ 13 & 14 & 5 \\ 2 & 10 & 4 \end{bmatrix}$$

ii) According to the definition of covariance and standard deviation, we have:

$$\text{corr}(\mathbf{u}, \mathbf{v}) = \frac{\text{cov}(\mathbf{u}, \mathbf{v})}{\sigma_{\mathbf{u}} \sigma_{\mathbf{v}}} = \frac{\frac{1}{n} \sum_{i=1}^n u_i v_i - \left(\frac{1}{n} \sum_{i=1}^n u_i\right) \left(\frac{1}{n} \sum_{i=1}^n v_i\right)}{\sqrt{\frac{1}{n} \sum_{i=1}^n u_i^2 - \left(\frac{1}{n} \sum_{i=1}^n u_i\right)^2} \sqrt{\frac{1}{n} \sum_{i=1}^n v_i^2 - \left(\frac{1}{n} \sum_{i=1}^n v_i\right)^2}}$$

Since \mathbf{u} and \mathbf{v} both have zero elementwise mean, then:

$$\text{corr}(\mathbf{u}, \mathbf{v}) = \frac{\text{cov}(\mathbf{u}, \mathbf{v})}{\sigma_{\mathbf{u}} \sigma_{\mathbf{v}}} = \frac{\frac{1}{n} \sum_{i=1}^n u_i v_i}{\sqrt{\frac{1}{n} \sum_{i=1}^n u_i^2} \sqrt{\frac{1}{n} \sum_{i=1}^n v_i^2}} = \frac{\sum_{i=1}^n u_i v_i}{\sqrt{\sum_{i=1}^n u_i^2} \sqrt{\sum_{i=1}^n v_i^2}} = \frac{|\mathbf{u}| |\mathbf{v}| \cos \theta}{|\mathbf{u}| |\mathbf{v}|} = \cos \theta.$$

So the correlation between the elements of \mathbf{u} and \mathbf{v} is equal to their cosine similarity.

4.2 Singular Value Decomposition

i) Since $M = U\Sigma V^T$, U has the same number of rows as that of M , which is m and V^T has the same number of columns as that of M , which is n . So V has n rows. Since U and V are orthogonal matrices, they are both square matrices. So the dimension of U is $m \times m$ and the dimension of V is $n \times n$.

According to the definition of matrix multiplication, the dimension of Σ is $m \times n$ to make the multiplication between U and Σ , Σ and V^T possible.

ii) Since U and V are orthogonal matrices, the product of each matrix with its transpose are identity matrices, which means, $UU^T = I_m$, $VV^T = I_n$.

If matrix M is invertible, which means $m = n = \text{rank}(M)$, the inverse of M is $M^{-1} = V\Sigma^{-1}U^T$. Here Σ is a symmetric diagonal matrices and all its diagonal entries are non-zero, so we could get Σ^{-1} by replacing all the diagonal entries of Σ with their reciprocal and have $\Sigma\Sigma^{-1} = I_m$. So $MM^{-1} = U\Sigma V^T V\Sigma^{-1}U^T = U\Sigma(V^T V)\Sigma^{-1}U^T = U(\Sigma\Sigma^{-1})U^T = UU^T = I_m$, which suggests that $M^{-1} = V\Sigma^{-1}U^T$ is the inverse matrix of M if M is invertible.

On the other hand, if matrix is not invertible, we could use SVD to get the pseudoinverse in the similar way: $M^+ = V\Sigma^+U^T$. The difference here is that: m may not equal to n and $\min(m, n)$ may not equal to the rank of M . Σ^+ is the pseudoinverse of Σ , which could be calculated by replacing every non-zero diagonal entry in Σ by its reciprocal and transposing the resulting matrix. So $\Sigma\Sigma^+$ is a $m \times m$ diagonal matrix and all its non-zero diagonal entries are 1. Then $MM^+ = U\Sigma V^T V\Sigma^+U^T = U(\Sigma\Sigma^+)U^T$, which is also a $m \times m$ diagonal matrix and all its non-zero diagonal entries are 1.

???再改改,反之是 $n \times n$, 假定 $m > n > \text{rank}(M) = k$.