Stats 315a: Statistical Learning
Problem Set 1

Dong-Bang Tsai*

Department of Applied Physics, Stanford University, Stanford, California 94305, USA

(Dated: January 24, 2012)

Problem 1

(a)

Problem 2 (ESL 2.4)

The squared distance from any sample point to the origin has a χ_p^2 distribution with mean p; therefore, since a prediction point x_0 is drawn from this distribution, it will have a expected squared distance p from the origin.

Because $z_i = a^T x_i$, and a is independent from x_i, we can conclude that $z_i \sim N$, normal distribution. Now, let's calculate the expectation value of z_i. We know that for a p dimensional vector a

$$E(z) = E(a^T x) = a^T E(x) = 0 \tag{1}$$

The co-variant can be given by

$$Cov(a^T x) = a^T a = 1 \tag{2}$$

As a result, we get $z_i \sim N(0, 1)$, and the expected squared distance will be $E(z^2) = 1$.

---

*Electronic address: dbtsai@stanford.edu

Problem 3

(a)

(b)

(c)

(d)

Problem 4

(a)

Problem 5

(a)

Problem 6

(a)

(b)

In the ridge regression approach, the equation we are trying to solve is

$$(X^T X + \lambda I)\beta = X^T y \tag{3}$$

The solution can be given by $\hat{\beta}_\lambda = (X^T X + \lambda I)^{-1} X^T y$. We add positive constant to diagonal of $X^T X$; therefore, $(X^T X + \lambda I)^{-1}$ is non-singular, even if $X^T X$ is not of full rank. As a result, the solution always exists, and is unique.

(c)

(d)

Using $X = UDV^T$

$$\begin{aligned}
\hat{\beta}_\lambda &= (X^T X + \lambda I)^{-1} X^T y \\
&= \left(V D^T U^T U D V^T + \lambda I\right) V D U^T y \\
&= V \left(D^2 + \lambda I\right) V^T D U^T y \\
&= V \left(D^2 + \lambda I\right) D U^T y
\end{aligned} \tag{4}$$