参考资料：

- 期望、方差、协方差及相关系数的基本运算 (https://blog.csdn.net/MissXy_/article/details/80705828)
- 高斯分布期望和方差的推导 (https://blog.csdn.net/su_jz/article/details/52579723)
- LR正则化与数据先验分布的关系？ (https://www.zhihu.com/question/23536142/answer/90135994)

\documentclass[fleqn]{article}

\usepackage[fleqn]{amsmath}

2 给定目标函数，求梯度

$f_1(w) = \|Xw - Y\|_2^2 + \lambda\|w\|_1^1$

$f_2(w) = \|Xw - Y\|_2^2 + \lambda\|w\|_2^2$

其中 $X \in \Re^{m\times n}, Y \in \Re^{m\times 1}, w \in \Re^{n\times 1}, |w\|_p^p = |w_1|^p + |w_2|^p + \ldots + |w_n|^p$

答:

$$\|Xw - Y\|_2^2 = \Sigma_{i=1}^m |x_i w - y_i|^2 = \Sigma_{i=1}^m (\Sigma_{j=1}^n x_{ij} w_j - y_i)^2$$
$$=> \frac{\partial\|Xw - Y\|_2^2}{\partial(w_j)} = 2\Sigma_{i=1}^m (\Sigma_{j=1}^n x_{ij} w_j - y_i) x_{ij} = 2x_j^T (Xw - Y)$$

$$\lambda\|w\|_2^2 = \lambda\Sigma_{j=1}^n |w_j|^2 = \lambda\Sigma_{j=1}^n w_j^2$$
$$=> \frac{\partial\lambda\|w\|_2^2}{\partial(w_j)} = 2\lambda w_j$$

$$\lambda\|w\|_1^1 = \lambda\Sigma_{j=1}^n |w_j|$$
$$=> \frac{\partial\lambda\|w\|_1^1}{\partial(w_j)} = \lambda sign(w_j)$$

$$\frac{\partial f_2(w)}{\partial(w_j)} = \frac{\partial\|Xw - Y\|_2^2}{\partial(w_j)} + \frac{\partial\lambda\|w\|_2^2}{\partial(w_j)} = 2x_j^T (Xw - Y) + 2\lambda w_j$$
$$=> \frac{\partial f_2(w)}{\partial(w)} = 2X^T (Xw - Y) + 2\lambda w$$

$$\frac{\partial f_1(w)}{\partial(w_j)} = \frac{\partial\|Xw - Y\|_2^2}{\partial(w_j)} + \frac{\partial\lambda\|w\|_1^1}{\partial(w_j)} = 2x_j^T (Xw - Y) + \lambda sign(w_j)$$
$$=> \frac{\partial f_1(w)}{\partial(w)} = 2X^T (Xw - Y) + \lambda sign(w)$$

3 给定随机变量 X 服从高斯分布 $N(\mu, \sigma^2)$

μ 和 σ^2 未知，从 X 中抽出 n 个样本 $(x_1, x_2, \ldots, x_n)$

我们想用这 n 个样本估算均值 μ 和方差 σ^2

我们估算的均值记为 $\hat{\mu}$ ，方差记为 $\hat{\sigma}^2$ ，估算的随机变量 $\hat{X}$ 。

我们现在有如下四种方法来估计 $\hat{\mu}$ ， n_0 是一个常数:

$$
\begin{aligned}
&(1)\hat{\mu} = \frac{x_1 + x_2 + \ldots + x_n}{n} \\
&(2)\hat{\mu} = \frac{x_1 + x_2 + \ldots + x_n}{n+1} \\
&(3)\hat{\mu} = \frac{x_1 + x_2 + \ldots + x_n}{n+n_0} \\
&(4)\hat{\mu} = 0
\end{aligned}
$$

我们想估计偏差和方差:

- 偏差 $E(\hat{X} - \mu)$
- 方差 $\hat{\sigma}^2 = Var[\hat{X}]$

(a) 各个方法的偏差是多少

答:

$$
\begin{aligned}
&E(\hat{X} - \mu) = E(\hat{X}) - \mu = \hat{\mu} - \mu \\
&(1)E(\hat{X} - \mu) = \frac{1}{n}\Sigma_{i=1}^{n} x_i - \mu \\
&(2)E(\hat{X} - \mu) = \frac{1}{n+1}\Sigma_{i=1}^{n} x_i - \mu \\
&(3)E(\hat{X} - \mu) = \frac{1}{n+n_0}\Sigma_{i=1}^{n} x_i - \mu \\
&(4)E(\hat{X} - \mu) = -\mu
\end{aligned}
$$

(b) 各个方法的方差是多少

答:

$$
\begin{aligned}
\hat{\sigma}^2 &= Var[\hat{X}] = E[(\hat{x}-\hat{\mu})^2] \\
E[(\hat{x}-\hat{\mu})^2] &= \int_{-\infty}^{+\infty} (\hat{x}-\hat{\mu})^2 \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-\frac{(\hat{x}-\hat{\mu})^2}{2\hat{\sigma}^2}} d\hat{x} \\
&= \int_{-\infty}^{+\infty} \hat{x}^2 \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-\frac{\hat{x}^2}{2\hat{\sigma}^2}} d\hat{x} \\
&= \hat{\sigma}\sqrt{2} \int_{-\infty}^{+\infty} (\hat{\sigma}\sqrt{2}\hat{x})^2 \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-\frac{(\hat{\sigma}\sqrt{2}\hat{x})^2}{2\hat{\sigma}^2}} d\hat{x} \\
&= \frac{4\hat{\sigma}^2}{\sqrt{\pi}} \int_{0}^{+\infty} \hat{x}^2 e^{-\hat{x}^2} d\hat{x} \\
let \quad \hat{x} &= t^2, => d\hat{x} = 2\sqrt{t}^{-1} dt \\
E[(\hat{x}-\hat{\mu})^2] &= \frac{4\hat{\sigma}^2}{\sqrt{\pi}} \int_{0}^{+\infty} t e^{-t} 2\sqrt{t}^{-1} dt \\
&= \frac{4\hat{\sigma}^2}{\sqrt{\pi}} \frac{1}{2} \int_{0}^{+\infty} t^{\frac{3}{2}-1} e^{-t} dt \\
&= \frac{4\hat{\sigma}^2}{\sqrt{\pi}} \frac{1}{2} \Gamma(\frac{3}{2}) \\
&= \frac{4\hat{\sigma}^2}{\sqrt{\pi}} \frac{1}{2} \frac{\sqrt{\pi}}{2} \\
&= \hat{\sigma}^2
\end{aligned}
$$

即 $\hat{\sigma}^2 = Var[\hat{X}] = E[(\hat{X}-\hat{\mu})^2] = \hat{\sigma}^2$

$\hat{\sigma}^2$ 与 $\hat{\mu}$ 的估计值无关。

(c) 计算期望

对 $\hat{X}$ 的估计是依赖样本 $(x_1, x_2, \dots, x_n)$ 的，假设我们从 $X' \sim N(\mu, \sigma^2)$ 中采到了新样本 x' ，X' 和 X 独立同分布，试求（用 $\hat{\mu}, \hat{\sigma}^2, \mu, \sigma^2$ 表达）$E[(\hat{X}-X')^2]$ 和 $E[(\hat{X}-\mu)^2]$ 的表达式。

答:

由 $\hat{\sigma}^2 = E[(\hat{X}-\hat{\mu})^2] = E(\hat{X}^2) - E(\hat{X})^2 = E(\hat{X}^2) - \hat{\mu}^2$

得 $E(\hat{X}^2) = \hat{\sigma}^2 + \hat{\mu}^2, E(X'^2) = \sigma^2 + \mu^2$

因为 X' 和 $\hat{X}$ 相互独立，所以 $E(\hat{X}X') = E(\hat{X})E(X') = \hat{\mu}\mu$

$$
\begin{aligned}
&E[(\hat{X} - X')^2] \\
&= E(\hat{X}^2) - E(2\hat{X}X') + E(X'^2) \\
&= \hat{\sigma}^2 + \hat{\mu}^2 - 2\hat{\mu}\mu + \sigma^2 + \mu^2 \\
&= \hat{\sigma}^2 + \sigma^2 + (\hat{\mu} - \mu)^2
\end{aligned}
$$

$$
\begin{aligned}
&E[(\hat{X} - \mu)^2] \\
&= E[((\hat{X} - \hat{\mu}) + (\hat{\mu} - \mu))^2] \\
&= E[(\hat{X} - \hat{\mu})^2] + 2(\hat{\mu} - \mu)E(\hat{X} - \hat{\mu}) + (\hat{\mu} - \mu)^2 \\
&= \hat{\sigma}^2 + (\hat{\mu} - \mu)^2
\end{aligned}
$$

(d) $E[(\hat{X} - \mu)^2]$ 被称为总误差 依次写出四种方法的总误差

答:

$$
\begin{aligned}
&E[(\hat{X} - \mu)^2] = \hat{\sigma}^2 + (\hat{\mu} - \mu)^2 = \hat{\sigma}^2 + (\frac{1}{n+n_0}\Sigma_{i=1}^n x_i - \mu)^2 \\
&(1) n_0 = 0, E[(\hat{X} - \mu)^2] = \hat{\sigma}^2 + (\frac{1}{n}\Sigma_{i=1}^n x_i - \mu)^2 \\
&(2) n_0 = 1, E[(\hat{X} - \mu)^2] = \hat{\sigma}^2 + (\frac{1}{n+1}\Sigma_{i=1}^n x_i - \mu)^2 \\
&(3) n_0 = n_0, E[(\hat{X} - \mu)^2] = \hat{\sigma}^2 + (\frac{1}{n+n_0}\Sigma_{i=1}^n x_i - \mu)^2 \\
&(3) n_0 = 0, E[(\hat{X} - \mu)^2] = \hat{\sigma}^2 + \mu^2
\end{aligned}
$$

(e) 判断偏差和方差的变化趋势

这四种估计方法都是第(3)种的特例。

- $n_0 = 0$ 时，得到 (1)
- $n_0 = 1$ 时，得到 (2)
- $n_0 = +\infty$ 时，得到 (4)

当 n_0 从 0 增加时，偏差和方差如何变化。

答：偏差 $E(\hat{X} - \mu) = \frac{1}{n+n_0}\Sigma_{i=1}^n x_i - \mu$ ，当 n_0 增加时，偏差减小。

方差 $\hat{\sigma}^2$ 与 n_0 无关，当 n_0 增加时，方差不变。

(f) 计算使总误差最小的参数

设 $n_0 = \alpha n$ ，如果 μ 和 σ^2 已知，当 α 为多少时，总误差最小?

答:

$$E[(\hat{X}-\mu)^2] = \hat{\sigma}^2 + (\hat{\mu}-\mu)^2 = \hat{\sigma}^2 + (\frac{1}{n(1+\alpha)}\Sigma_{i=1}^n x_i - \mu)^2$$

$$let \quad \frac{\partial E[(\hat{X}-\mu)^2]}{\partial \alpha} = 2(\frac{1}{n(1+\alpha)}\Sigma_{i=1}^n x_i - \mu)(\Sigma_{i=1}^n x_i)(-\frac{n}{(n(1+\alpha))^2}) = 0$$

$$=> \frac{1}{n(1+\alpha)}\Sigma_{i=1}^n x_i - \mu = 0$$

$$=> \alpha = \frac{1}{n\mu}\Sigma_{i=1}^n x_i - 1$$

(g) 最大后验计算

用样本 $(x_1, x_2, \ldots, x_n)$ 进行 $\hat{\mu}$ 的估计，设假设集合 $H = \left\{\hat{\mu}; \hat{\mu} \in \mathfrak{R}\right\}$ ，采用高斯先验 $p(\mu) = \frac{1}{\bar{\sigma}\sqrt{2\pi}}e^{-\frac{\mu^2}{2\bar{\sigma}^2}}$ 。

- 1) 导出 $\hat{\mu}$ 的估计式；
- 2) 尝试建立参数 $\bar{\sigma}^2$ 和(f)中的 α 的关系；
- 3) 试建立参数α 和正则系数 λ 的关系。

1) 答:

$$\begin{aligned} argmax P(\hat{\mu}|\hat{x}) &= argmax P(\hat{x}|\hat{\mu})P(\hat{\mu})/P(\hat{x}) \\ &= argmax P(\hat{x}|\hat{\mu})P(\hat{\mu}) \\ &= argmzx(\frac{1}{\hat{\sigma}\sqrt{2\pi}}e^{-\frac{(\hat{x}-\hat{\mu})^2}{2\hat{\sigma}^2}}\frac{1}{\bar{\sigma}\sqrt{2\pi}}e^{-\frac{\hat{\mu}^2}{2\bar{\sigma}^2}}) \end{aligned}$$

$$argmin(NLL) = argmin(log(\hat{\sigma}\bar{\sigma}2\pi) + \frac{(\hat{x}-\hat{\mu})^2}{2\hat{\sigma}^2} + \frac{\hat{\mu}^2}{2\bar{\sigma}^2})$$

$$\begin{aligned} let \quad \frac{\partial NLL}{\partial \hat{\mu}} &= -\frac{2(\hat{x}-\hat{\mu})}{2\hat{\sigma}^2} + \frac{2\hat{\mu}}{2\bar{\sigma}^2} = 0 \\ => \hat{\mu} &= \frac{\bar{\sigma}^2}{\hat{\sigma}^2+\bar{\sigma}^2}\hat{x} \end{aligned}$$

$$\begin{aligned} use \quad \hat{x} &= (x_1, x_2, \ldots, x_n) \quad to\ add \\ => n\hat{\mu} &= \frac{\bar{\sigma}^2}{\hat{\sigma}^2+\bar{\sigma}^2}\Sigma_{i=1}^n x_i \\ => \hat{\mu} &= \frac{\bar{\sigma}^2}{\hat{\sigma}^2+\bar{\sigma}^2}\frac{1}{n}\Sigma_{i=1}^n x_i \end{aligned}$$

2) 答:

$$for\ \ 2)\quad \hat{\mu} = \frac{\bar{\sigma}^2}{\hat{\sigma}^2 + \bar{\sigma}^2}\frac{1}{n}\Sigma_{i=1}^n x_i$$
$$and\ (f)\quad n_0 = \alpha n,\ i.e.\ \hat{\mu} = \frac{1}{1+\alpha}\frac{1}{n}\Sigma_{i=1}^n x_i$$

$$=> \frac{1}{1+\alpha} = \frac{\bar{\sigma}^2}{\hat{\sigma}^2 + \bar{\sigma}^2}$$
$$=> \bar{\sigma}^2 = \frac{1}{\alpha}\hat{\sigma}^2$$

3) 答:

真实值 $\mu = \hat{x}w + \epsilon,\ \epsilon \sim N(0, \tilde{\sigma})$

估计值 $\hat{\mu} = \hat{x}w$，$\hat{X} = (x_1, x_2, \ldots, x_n)$，$w = \frac{1}{n+n_0}(1, 1, \ldots, 1)$

$$\begin{aligned}
P(\mu) &= P(\mu|\hat{X}, \hat{\mu})P(\hat{X}|\hat{\mu})P(\hat{\mu}) \\
&= \frac{1}{\tilde{\sigma}\sqrt{2\pi}}e^{-\frac{(\mu-\hat{x}w)^2}{2\tilde{\sigma}^2}}\frac{1}{\hat{\sigma}\sqrt{2\pi}}e^{-\frac{(\hat{x}-\hat{\mu})^2}{2\hat{\sigma}^2}}\frac{1}{\bar{\sigma}\sqrt{2\pi}}e^{-\frac{\hat{\mu}^2}{2\bar{\sigma}^2}} \\
NLL &= C' + \frac{(\mu - \hat{x}w)^2}{2\tilde{\sigma}^2} + \frac{(\hat{x} - \hat{\mu})^2}{2\hat{\sigma}^2} + \frac{\hat{\mu}^2}{2\bar{\sigma}^2} \\
&= C' + \frac{(\mu - \hat{x}w)^2}{2\tilde{\sigma}^2} + \frac{\hat{x}^2(1-w)^2}{2\hat{\sigma}^2} + \frac{\hat{x}^2w^2}{2\bar{\sigma}^2} \\
&= C' + \frac{(\mu - \hat{x}w)^2}{2\tilde{\sigma}^2} + \frac{\hat{x}^2}{2\hat{\sigma}^2} - \frac{\hat{x}^2}{\hat{\sigma}^2}w + (\frac{1}{2\hat{\sigma}^2} + \frac{1}{2\bar{\sigma}^2})\hat{x}^2w^2
\end{aligned}$$

$$last\ items\ is\quad \lambda w^2,\ \ and\ for\quad \bar{\sigma}^2 = \frac{1}{\alpha}\hat{\sigma}^2$$
$$=> \lambda = (\frac{1}{2\hat{\sigma}^2} + \frac{1}{2\bar{\sigma}^2})\hat{x}^2 = \frac{1+\alpha}{2\hat{\sigma}^2}\hat{x}^2$$
$$use\quad \hat{x} = (x_1, x_2, \ldots, x_n)\quad to\ add$$
$$=> n\lambda = \frac{1+\alpha}{2\hat{\sigma}^2}\Sigma_{i=1}^n x_i^2$$
$$=> \lambda = \frac{1+\alpha}{2\hat{\sigma}^2}\frac{1}{n}\Sigma_{i=1}^n x_i^2$$

(h) 开放式问题

如果我们已知真实的 μ 大概分布在某个值 μ_0 附近，我们应该如何估计 $\hat{\mu}$，这对你用 MAP / L2 正则的选取有何启示?

In [ ]:

```

```