1 Формулировка задачи

1.1

$\widehat{y} = Xw$ - каждое значение целевой переменной есть сумма признака на соответствующий вес

1.2

$\langle y - \widehat{y} \rangle \langle y - \widehat{y} \rangle = \langle y - Xw \rangle \langle y - Xw \rangle = (y - Xw)^T (y - Xw)$

1.3

$\sum_{i=1}^{n} \omega_i x_i$ - линейная комбинация(x_i - столбцы, соответствующие i признаку)

2 Нормальное уравние

2.1

Вспомогательные утверждения: $\dfrac{\partial b^T b}{\partial b} = 2b$ - очевидно, $\dfrac{\partial Ax + b}{\partial x} = A^T$ - обозначим $f = Ax + b$, $\dfrac{\partial f}{\partial x} = \left(\dfrac{\partial f}{\partial x_1}, ..., \dfrac{\partial f}{\partial x_n} \right)^T$ получим A^T Также существует правило композиции как и для обычных функций из матана. Поэтому $\dfrac{\partial (Ax+b)^T (Ax+b)}{\partial x} = 2(Ax+b)A^T$

2.2

Воспользуемся равенством выше, получим $2X^T(X\omega - y) = 0$ в итоге, раскрывая скобки $\omega = (X^T X)^{-1} X^T y$

3 Геометрическая интерпретация

3.1

Данное выражение равно нулю. Раскроем скобки в $\langle x^j, y - \widehat{y} \rangle$: первое скалярное произведение $\langle x^j, y \rangle$ проекция y на x^j, второе скалярное произведение проекция проекции y на x^j по теореме о трёх перпендикулярах они равны.

3.2

Нулевому вектору. Так как каждая координата равна нулю по доказанному выше.

Подставляем выражение для $\widehat{y}$ получаем ответ: $\omega = (X^T X)^{-1} X^T y$

4 Вероятностная интрепретация

4.1

$\rho(x) = e\dfrac{(x-a)^2}{2\sigma^2}$ - одномерная плотность нормального распределения
$L = \prod_{i=1} e\dfrac{(x_i-a)^2}{2\sigma^2} \rightarrow max$, где L правдоподобие. Максимизировать это произведение сразу сложно, поэтому его логарфимируют, логарифм - функция монотонная, поэтому сохраняет экстремумы. В итоге получаем выражение - $\sum_{i=1}^{n} \dfrac{(x_i-a)^2}{2\sigma^2}$, которое нужно максимизировать

4.2

$\sum_{i=1}^{n} \dfrac{(x_i-a)^2}{2\sigma^2}$

4.3

Показано выше. Выносим дисперсию - в итоге нужно получить сумму квадратов отклонений.

4.4

Распределение Лапласа - $\rho(x) = \dfrac{\alpha}{2} e^{-\alpha|x-\beta|}$

5 Регуляризация

$\langle y - Xw\rangle\langle y - Xw\rangle = (y - Xw)^T(y - Xw) + \tau\omega^T\omega$. Дифференциируя и сокращая это выражение получим: $X^T(X\omega - y) + \tau\omega = 0 \Leftrightarrow X^T X\omega - X^T y + \tau E\omega \Leftrightarrow (X^T X + +\tau E)\omega = X^T y \rightarrow \omega = (X^T X + \tau E)^{-1} X^T y$

5.1

См.выше.

5.2

Собственное значение определяется как $\max_{\|u\|=1} \parallel \Sigma' u \parallel$, где $\Sigma' = X^T X + \tau E$ По неравеству Коши выполняется $\parallel \Sigma u \parallel \leq \parallel \Sigma \parallel \parallel u \parallel$ следовательно максимум равен этому значению $\parallel \Sigma \parallel \parallel u \parallel$, а оно ограничено по неравенству треугольника - $(\parallel \Sigma \parallel + \lambda \parallel E \parallel) \parallel u \parallel$, что равно $(\lambda + \tau)$. Покажем, что собственные векторы матриц $\Sigma^{'}$ и $\Sigma = X^T X$ - одинаковые $\Sigma^{'} u = (X^T X + \tau E) u = X^T X u + \tau E u = (\lambda + t) u$ - следовательно собственные векторы, остаются одними и теми же.

6 Лассо Тибширани

6.1

6.2

Рассмотрим значение градиента функции в неугловой точки, оно отлично от нуля следовательно можем сместиться в соседние точки и так для каждой неугловой точки.